Tandem Mass Spectrometry and Neonatal Blood Screening in Quebec

Summary Report
Tandem Mass Spectrometry and Neonatal Blood Screening in Quebec

Summary Report

Report prepared for AETMIS by
Héla Makni, Carole St-Hilaire, Laura Robb, Kathy Larouche and Ingeborg Blancquaert

November 2007
This summary was translated from an official French publication of the Agence d'évaluation des technologies et des modes d'intervention en santé (AETMIS). Both the original summary report, titled *La spectrométrie de masse en tandem et le dépistage néonatal sanguin au Québec : rapport sommaire* and the English summary are available in PDF format on the Agency’s Web site.

**SCIENTIFIC REVIEW**
Ingeborg Blancquaert, Scientific Advisor and Coordinator of the Genetics Module

**TRANSLATION**
Mark Wickens, PhD, Certified Translator

**PAGE LAYOUT**
Jocelyne Guillot

**BIBLIOGRAPHIC VERIFICATION**
Denis Santerre

**COORDINATION**
Lise-Ann Davignon

**COORDINATION OF EXTERNAL REVIEW**
Valérie Martin

**INFORMATION SPECIALISTS**
Mathieu Plamondon
Fleurette Grégoire

**COMMUNICATIONS AND DISSEMINATION**
Diane Guibault
Richard Lavoie

For further information about this publication or any other AETMIS activity, please contact:

Agence d'évaluation des technologies et des modes d'intervention en santé
2021, Union Avenue, Suite 10.083
Montréal (Québec) H3A 2S9

Telephone: 514-873-2563
Fax: 514-873-1369
E.mail: aetmis@aetmis.gouv.qc.ca
www.aetmis.gouv.qc.ca

How to cite this document:


Legal deposit
Bibliothèque et Archives nationales du Québec, 2007
Library and Archives Canada, 2007

This report may be reproduced in whole or in part provided that the source is cited.
The mission of the Agence d’évaluation des technologies et des modes d’intervention en santé (AETMIS) is to help improve the Québec health-care system. To this end, it advises and supports the Minister of Health and Social Services and decision-makers in the health-care system with regard to the assessment of health services and technologies. The Agency makes recommendations based on scientific reports assessing the introduction, diffusion and use of health technologies, including technical aids for the disabled, as well as the methods of providing and organizing services. The assessments examine many different factors, such as efficacy, safety and efficiency, as well as ethical, social, organizational and economic issues.

THE BOARD

Dr. Jeffrey Barkun, Surgeon, Royal Victoria Hospital, MUHC, and Director, Department of General Surgery, Faculty of Medicine, McGill University, Montréal

Dr. Marie-Dominique Beaulieu, Holder of the Dr. Sadok Besrou Chair in Family Medicine, CHUM, Full Professor, Faculty of Medicine, Université de Montréal, and Researcher, Evaluative Research Unit, Hôpital Notre-Dame, CHUM, Montréal

Dr. Sylvie Bernier, Director, Organization of Medical and Technological Services, MSSS, Québec

Dr. Suzanne Claveau, Microbiology and Infectious Disease Specialist, Hôtel-Dieu de Québec, CHUQ, and Associate Clinical Professor, Faculty of Medicine, Université Laval, Québec

Dr. Serge Dubé, Surgeon, Director of the Surgery Program, Hôpital Maisonneuve-Rosemont, and Vice-Dean of Professorial Affairs, Faculty of Medicine, Université de Montréal

Roger Jacob, Biomedical Engineer, Coordinator, Capital Assets and Medical Equipment, Agence de la santé et des services sociaux de Montréal

Dr. Michel Labrecque, Professor and Clinical Researcher, Family Medicine Unit, Hôpital Saint-François d’Assise (CHUQ), Québec

A.-Robert LeBlanc, Engineer, Full Professor and Program Director, Biomedical Engineering Institute, Université de Montréal, and Assistant Director of Research, Development and Utilization, Hôpital du Sacré-Cœur de Montréal Research Centre, Montréal

Esther Leclerc, Registered Nurse, Director of Nursing, Hôpital Saint-Luc, CHUM

Dr. Jean-Marie Moutquin, Obstetrician/Gynecologist, Director of Research and Director, Department of Obstetrics and Gynecology, CHUS, Sherbrooke

Dr. Réginald Nadeau, Cardiologist, Researcher, Hôpital du Sacré-Cœur de Montréal Research Centre, and Emeritus Professor, Faculty of Medicine, Université de Montréal

Johane Patenaude, Ethicist, Associate Professor, Department of Surgery, Faculty of Medicine, Université de Sherbrooke, and FRSQ Research Scientist

Dr. Simon Racine, Community Health Specialist, Deputy Chief Executive Officer, Clinical Affairs, Centre hospitalier Robert-Giffard – Institut universitaire en santé mentale, Québec

Lee Soderstrom, Economist, Associate Professor, Department of Economics, McGill University, Montréal
This brief report is a summary of a technical report prepared at the request of the Ministère de la Santé et des Services sociaux. Both reports stem from the same literature review and include the same sections, although these sections are much more detailed in the technical report, especially with regard to the review on the diseases of interest, the technical aspects of tandem mass spectrometry, and the different issues associated with the use of this technology.
Foreword

Tandem Mass Spectrometry and Neonatal Blood Screening in Quebec

This report was prepared at the request of the Ministère de la Santé et des Services sociaux (MSSS), in the context of scientific debates and pressure in favour of adopting tandem mass spectrometry (MS/MS) for neonatal blood screening of inborn errors of metabolism. MS/MS can be used to simultaneously screen for more than 30 inborn errors of metabolism in a single analytical step with a high throughput. In its request, the MSSS asked AETMIS to evaluate whether it would be pertinent to use MS/MS for neonatal blood screening in Quebec. Once the systematic reviews and the available Quebec data had been analysed, it was agreed that AETMIS would 1) examine the relevance of replacing the current screening methods for phenylketonuria (PKU) and tyrosinemia type 1 (TT1) by MS/MS and of introducing neonatal screening for Medium-Chain Acyl-CoA Dehydrogenase Deficiency (MCADD); and 2) analyze the main ethical, social, economic and organizational issues. The expansion of neonatal screening to other diseases could possibly be the subject of a subsequent report.

Our review confirms the importance of a case-by-case analysis for each inborn error of metabolism. Indeed, the available options depend on the specific characteristics and state of knowledge for each disease, and the applicability of the technological developments to these diseases. Even though there are gaps in the data, current evidence supports the clinical utility of neonatal screening for the three diseases in question. As for the appropriateness of implementing MS/MS-based screening in Quebec, the situation differs according to the disease. For MCADD, MS/MS is the only technology available for neonatal screening, and its performance is one of the best for this particular condition. For PKU, the literature suggests that MS/MS yields fewer false positives than the current technology, but compared to the results observed in Quebec, this advantage would not be substantial. However, if MS/MS were used for MCADD screening, the technology transfer for PKU would avoid a duplication of analytical steps and would be efficient, according to the health economics literature examined. For TT1, MS/MS-based neonatal screening relying on both tyrosine and succinylacetone assays seems promising but needs further validation. Furthermore, the judiciousness of a technology transfer and its optimal timing depend on a number of ethical, social, legal, economic and organizational issues, in addition to the scientific and technical considerations. Therefore, three separate scenarios are proposed for consideration by policy-makers:

1) conducting a pilot study;
2) postponing the introduction of MS/MS until after the necessary validation studies for TT1 screening have been completed; and
3) introducing MS/MS for PKU and MCADD screening, while, either undertaking gradual technology replacement for TT1, or maintaining the current methods until the results of the validation studies are available.

Whichever option is chosen, implementing MS/MS must not be done hastily, since other issues—ethical, economic and organizational—first need to be resolved.

In submitting this report, AETMIS wishes to contribute to decision-making regarding policies governing Quebec’s neonatal blood screening program.

Dr. Juan Roberto Iglesias
President and Chief Executive Officer
This report was prepared at the request of the Agence d’évaluation des technologies et des modes d’intervention en santé (AETMIS) by Dr. Héla Makni, M.Sc. (Epidemiology and Biostatistics), Carole St-Hilaire, Economist, Ph.D. (Public Health), Laura Robb, M.Sc. (Genetic Counselling), Kathy Larouche, M.Sc. (Physiology and Endocrinology), and Dr. Ingeborg Blancquaert, Pediatrician, Ph.D. (Epidemiology), Coordinator of the Genetics Module and Scientific Advisor, all of whom are research consultants at AETMIS.

AETMIS cordially thanks the external reviewers for their valuable suggestions and their contribution to the overall quality and the rigour of this assessment report:

**Pierre Allard**
Biochemist, Department of Medical Genetics, Centre hospitalier universitaire Saint-Justine, Montréal, Quebec

**Denise Avard**
Researcher, Public Law Research Centre, Faculty of Law, Université de Montréal, Montréal, Quebec

**Dr. Ségolène Aymé**
Research Director, Institut national de la santé et de la recherche médicale (INSERM), Paris, France

**Dr. David Cheillan**
*Praticien hospitalier*, Neonatal Screening Unit, Department of Pediatric Biochemistry, Hôpital Debrousse, Lyon, France

**Dr. Régen Drouin**
Geneticist, Department of Genetics, and Associate Professor, Department of Pediatrics, Faculty of Medicine and Health Sciences, Université de Sherbrooke, Sherbrooke, Quebec

**Dr. Jean-Pierre Farriaux**
Honorary Professor of Pediatric Medical Genetics, Centre hospitalo-universitaire de Lille, Association française pour le dépistage et la prévention des handicaps de l’enfant (AFDPHE), France

**Scott Grosse**
Senior Health Economist, National Center on Birth Defects and Developmental Disabilities, Centers for Disease Control and Prevention, Atlanta, GA, United States

**Dr. Serge Melançon**
Director, Medical Genetics Clinic, Montreal Children’s Hospital, and Director, Department of Biochemical Genetics, McGill University Health Centre, Montréal, Quebec
The Agency also wishes to express its gratitude to the following experts, whose contribution enriched the contents of this report:

**Jean-Marie Lance**, M.Sc. (Economics), Senior Scientific Advisor, AETMIS, and **Lee Soderström**, Ph.D. (Economics), Associate Professor, Department of Economics, McGill University, who reviewed the economic section of this report;

**Ghislaine Cleret de Langavant**, Ph.D. (Bioethics), Research Consultant, AETMIS, who helped draft and revise certain parts of this report;

Dr. **Bridget Wilcken**, Professor, NSW Biochemical Genetics and Newborn Screening Service, Children’s Hospital, Westmead, Sydney, Australia, and Dr. **François Rousseau**, Full Professor, Department of Medical Biology, Faculty of Medicine, Université Laval, Quebec City, who kindly provided us with unpublished data;

**Pierre Allard** and **Raymond Lepage**, Department of Clinical Biochemistry, Centre hospitalier de l’Université de Montréal; **Robert Giguère**, Department of Genetics, Centre hospitalier de l’Université de Sherbrooke; and **Jean Ruel**, laboratory technician for the screening program at the Centre hospitalier universitaire de Québec, who provided us with information on tandem mass spectrometry and other technologies used for neonatal screening of inborn errors of metabolism.

**CONFLICT OF INTEREST**

None declared.
In Quebec, three diseases—phenylketonuria (PKU), tyrosinemia type 1 (TT1) and congenital hypothyroidism—are presently screened for in neonates. With the development of tandem mass spectrometry (MS/MS), the scientific community engaged in a debate over the relevance of expanding neonatal screening programs to include a number of inborn errors of amino acid, fatty acid and organic acid metabolism. Pressure in this regard is being exerted by health professionals, patient groups and industry. This technology can be used to selectively screen for specific inborn errors of metabolism or to perform a full scan of the metabolic profiles associated with more than 30 such diseases. Two separate decisional questions need to be addressed: 1) whether MS/MS should replace the technologies currently used for PKU and TT1 screening, and 2) whether neonatal screening should be expanded to other diseases. The second issue requires rigorous evaluation of the clinical utility and relevance of screening for each inborn error of metabolism. According to the literature, the disease that ranks highest as a potential candidate for inclusion in neonatal screening is Medium-Chain Acyl-CoA Dehydrogenase Deficiency (MCADD).

The advantages of the MS/MS technology highlighted by its promoters are its ability to screen for a large number of diseases in a single analytical step, its ability to screen for diseases, such as MCADD, that are not detectable by other techniques, the fact that the technology is automated and thus suitable for high throughput and the fact that its performance is better than that of other screening techniques (e.g., fewer false positives for PKU). While recognizing these theoretical advantages, the authors of systematic reviews published thus far have pointed out several significant knowledge gaps which decrease the weight of these assertions. They specifically state that the evaluation of MS/MS performance is mainly based on studies providing results for groups of inborn errors of metabolism and that few studies have evaluated this performance for individual diseases. Given the study designs used, uncertainty persists as to the proportion of false negative results. These estimates depend on the manner in which clinical services are organized for rare metabolic diseases. In addition, for a number of diseases, there is a paucity of epidemiological data, the natural course of the disease is not well known and the clinical benefits of screening have not been clearly demonstrated. Some of these limitations are due to the difficulty in gathering data on rare diseases, for which establishing a clinical diagnosis can be complex.

Given these limitations, the authors of systematic reviews were cautious in their recommendations. The first reviews, which date from 1997, called attention to the technology’s potential but did not recommend technology replacement. A number of authors recommended instituting large-scale pilot projects for PKU and MCADD in order to evaluate the performance of MS/MS screening, its efficiency and the impact on health outcomes. A review of the recent literature shows that more studies are presenting MS/MS performance data for the screening of individual diseases. However, no new studies specific to MCADD screening were identified. Even if preliminary data on the impact of early diagnosis and management for MCADD following screening are starting to accumulate, this literature is based on a limited number of cases with

1. MS/MS can not be used for congenital hypothyroidism screening.
2. PKU and TT1 are inborn errors of amino acid metabolism, while MCADD is a disorder of fatty acid metabolism.
3. One of the two British reports adds glutaric acidemia type I to this list [Seymour et al., 1997].
insufficient follow-up and no prospective, controlled studies have been published. The most recent systematic reviews nevertheless tend to recommend the implementation of PKU and MCADD screening programs. These recommendations are supported by the results of several economic modelling exercises suggesting that MS/MS screening is efficient if performed for at least two diseases, including PKU. It seems that these shifting views are not so much due to the accumulation of new knowledge as to the difficulty in implementing the previously recommended pilot projects. The above-mentioned caution, which limits the recommendations to screening for a few diseases, is by no means generalized and a number of non-systematic reviews and publications, based mainly on expert opinions, have come out in favour of technology replacement and expansion of neonatal screening to a wide range of inborn errors of metabolism. These recommendations have led to decisions along these lines in several jurisdictions, but have also prompted a debate over the role of evidence in such decisions.

This report examines the relevance of replacing current technologies by MS/MS for PKU and TT1 screening and of including MCADD in Quebec’s neonatal blood screening program. It is based on a review of the literature on the natural course of the three diseases of interest, their epidemiology and the efficacy of therapeutic measures, the performance of MS/MS technology, and cost and efficiency data. A cost analysis and a review of the ethical, psychosocial and organizational issues round out the report. The issue of expanding screening to include other diseases could be considered in a subsequent report, with candidate diseases prioritized according to the available data.

Our review of the evidence on MS/MS performance confirms the reservations expressed in the previous systematic reviews with respect to the quality of available studies. These reservations specifically concern study design, reporting of results, study population selection processes and the lack of standardization with regard to several factors which could affect the quality of MS/MS analyses and to the diagnostic confirmation tests. Most of the available evidence derives from prospective cohort studies carried out in the context of neonatal screening programs because of the difficulty in conducting comparative prospective studies with a suitable control group, as pointed out by several authors. In newborn screening programs, the reference tests to confirm the diagnosis are only performed for patients with a positive MS/MS result. Consequently, the data regarding the proportion of false negatives is of uncertain quality.

Overall, the results of the different studies that were reviewed indicate that the sensitivity, negative predictive value and specificity of MS/MS are high, whether for the neonatal screening of groups of diseases or for the selective screening of PKU, TT1 or MCADD. It is nonetheless possible that the sensitivity and negative predictive value were overestimated because of the quality of the data on false negatives. Furthermore, a considerable variability was observed for positive predictive values, even though there were only minor differences in the prevalence of inborn errors of metabolism between studies. The heterogeneity in the study populations’ characteristics, the age of sampling, the choice of metabolic markers, the cut-off values, the protocols for classifying MS/MS results and the diagnostic confirmation tests can influence MS/MS specificity and explain the variability in positive predictive values.

These performance data would not be automatically applicable if different analytical protocols than those used in the reviewed studies were implemented. Even today, the technology is constantly evolving in terms of the analytical procedures. Some of the technological changes on the horizon that could significantly alter MS/MS performance include incorporating the analysis of other metabolites (succinylacetone) into the protocol presently used for TT1 screening and eliminating the derivatization
process from sample preparation. The performance of these new approaches will need to be evaluated rigorously prior to implementation. In addition, the evolving nature of the technology underscores the importance of carefully considering technological options prior to implementation, performing analytical validation following any change to the protocols, and establishing ongoing quality assurance mechanisms.

The decision to include a given disease in a neonatal screening program is based, apart from considerations related to the technology’s performance, on the ability to favourably alter prognosis following early detection and intervention. With regards to the clinical utility for patients and their families, neonatal screening is justified for the three diseases of interest, despite gaps in the knowledge base and various issues raised for each disease. Our review confirms the importance of a case-by-case analysis for each disease of interest, since the available options depend on the specific characteristics and state of knowledge for each disease and on the applicability of the technological developments to these diseases.

- For MCADD, neonatal screening can only be carried out with MS/MS, the performance of which is particularly high for this condition. Knowledge of the entire spectrum of clinical forms is limited, especially for the less severe forms, and the variability in phenotypic expression makes it more difficult to compare the prognosis with and without screening and early management. The benefits of early treatment are convincing, however, for the severe end of the spectrum. Periodic reassessment of MCADD screening benefits, through ongoing data collection, will therefore be essential.

- For PKU, the literature suggests that MS/MS yields fewer false positives than the current technology, but compared to the results observed in Quebec, this advantage would not be substantial. According to the health economics literature, technology replacement is efficient if screening is carried out for at least two diseases, including PKU. If MS/MS were used for MCADD screening, the technology transfer for PKU would avoid duplication of analytical steps and would probably be less expensive than continuing with the current analytical method alongside MS/MS.

- For TT1, data supporting the efficacy of NTBC therapy\(^4\) are starting to accumulate, and they corroborate the utility of neonatal screening in Quebec. New approaches to TT1 screening based on assaying both tyrosine and succinylacetone seem promising but need further validation.

The review of the economic literature is aimed at documenting the costs, cost-effectiveness and cost-utility of MS/MS-based neonatal blood screening. This literature is characterized by wide differences in the inborn errors of metabolism considered, as well as by variability in incidence data, in probabilities of neurological disabilities and death in unscreened MCADD children and in the measures of efficacy used\(^5\). All studies tend to support the efficiency of MS/MS-based screening for several inborn errors of metabolism, especially those that cannot be screened for otherwise, as is the case with MCADD. In fact, all studies show that MS/MS is efficient if at least two diseases are screened for, including PKU. Lastly, it is noteworthy that no economic assessment has specifically examined the efficiency of MS/MS-based screening for TT1.

\(^{4}\) NTBC [2-(2-nitro-4-trifluoromethyl-benzoyl)-cyclohexane-1,3-dione] therapy is available under the name of nitisinone (Orfadin\(^b\)).

\(^{5}\) Outcomes considered for evaluating efficacy may include medical complications, hospitalizations, care and treatment, moderate or severe neurological disabilities, and prevented deaths.
The economic section of this report also provides budgetary information on certain capital and operating costs relevant to this type of screening. Thus, a budget impact approach was used to estimate the cost of MS/MS-based screening for PKU, TT1 and MCADD. Equivalent annual incremental costs (EAIC) were estimated. These allow the costs for the MS/MS and ancillary equipment, as well as for laboratory facility installation, to be spread out over several years. The EAIC thus represent the annual value of resources used for MS/MS-based neonatal blood screening. The results show that, in the Quebec context, this screening program would cost approximately CA$255,231 annually for a single neonatal screening laboratory. The main expenditures would be for acquisition of the MS/MS and ancillary equipment, and for laboratory technicians. Cost variations according to the equipment lifespan, the type of maintenance plan and the number of full-time equivalent technicians are presented. Lastly, it should be noted that cost estimations do not include expenditures associated with sample collection, interpretation of results, diagnostic confirmation, follow-up of patients and families and database management. These expenditures are assumed to be the same as with current screening methods.

The relevance and optimal timing of implementing MS/MS-based screening in Quebec depend on a number of ethical, social, legal, economic and organizational issues, in addition to scientific and technical considerations. Some of these issues are discussed here, while others are beyond the scope of this report. Three separate scenarios are therefore proposed for consideration by policy-makers:

1) Conducting a pilot study on the screening of these three diseases over several years.

2) Postponing MS/MS implementation until after validation studies for succinylacetone assays have been completed and then implementing a single analytical protocol for neonatal screening of the three diseases.

3) Introducing MS/MS-based screening for PKU and MCADD while, either undertaking gradual technology replacement for TT1, or maintaining the current methods until the results of the validation studies are available.

Each scenario has its pros and cons and entails different repercussions, both in terms of service organization and access to care. The choice between these three options is, of course, based on value judgments, but also depends on more concrete issues. The latter relate to the time required to prepare for technology implementation and to conduct—in Quebec or elsewhere—validation studies of the simplified protocol for TT1 screening, as well as to anticipated difficulties with a phased-in MS/MS implementation. The decision as to the best time to implement MS/MS will involve a trade-off between favouring a rapid access to services and opting to introduce the technology on the basis of data that are scientifically sound and/or applicable to Quebec. More thorough data on the recently developed analytical protocols for TT1 screening could be derived from a validation study and Quebec could be a favourable environment to conduct such a study. As for the advantages of a pilot study, these include gathering epidemiological and genetic data on MCADD and evaluating the costs that are directly applicable in Quebec. Such a pilot project however, is not likely to provide, within a reasonable timeframe, the data required for a definite evaluation of the benefits of MCADD screening in terms of long-term prognosis. It will therefore be essential to periodically reassess the benefits of neonatal MCADD screening.

Whichever option is chosen, implementing MS/MS must not be done hastily, since other issues need to be resolved beforehand. The policy regarding implicit consent for neonatal screening needs to be reviewed, particularly if the decision is taken...
to add a new disease to the screening program. Indeed, the procedure adopted in Quebec to justify inclusion of neonatal screening in routine care will pose problems in that event. There needs to be a consensus on practices following detection of non-targeted inborn errors of metabolism and on the protocol for MCADD diagnostic confirmation. A more thorough analysis of the feasibility of implementing MS/MS must also be carried out, taking into consideration, amongst other things, the capital and operating costs for each of the above-mentioned scenarios. At each stage of implementation, organizational issues must be addressed in order to prospectively optimize practices and generate the data needed to monitor program performance and to periodically evaluate the pertinence of choices made.

Lastly, it is worth mentioning the frequently raised concerns about the use of MS/MS technology to scan for complete metabolic profiles detecting more than 30 inborn errors of metabolism. Once the MS/MS technology is implemented, there will be an increased pressure to expand the neonatal screening program to several other inborn errors of metabolism. This pressure will be exerted by health professionals and industry, as well as by parent associations and the general public, who are increasingly informed through Internet. The arguments fuelling this pressure include the minimal costs of adding other inborn errors of metabolism once the technology is in place, the advantage of gathering data for research, the benefits for families and the ability to capitalize on what is considered the main advantage of MS/MS, namely, its capacity to analyze several metabolites simultaneously. Under no circumstances should screening for additional diseases be considered without a prior evaluation of the evidence and criteria that should guide the implementation of population-based screening programs. Finally, several problems discussed in this report, particularly those concerning the provision of information to parents and the availability of an effective network for patient management and follow-up by competent professionals, must necessarily be assessed and solved prior to any expansion of neonatal screening to other inborn errors of metabolism.
ABBREVIATIONS AND ACRONYMS

ACIEM Advisory Committee on Inborn Errors of Metabolism
AETMIS Agence d’évaluation des technologies et des modes d’intervention en santé
AFDPHE Association française pour le dépistage et la prévention des handicaps de l’enfant
ARG Arginase deficiency
ASL Argininosuccinate lyase deficiency
ASS Argininosuccinate synthetase deficiency, or citrullinemia
β-KT Beta-ketothiolase deficiency
C2 Acetylcarnitine
C3 Propionylcarnitine
C4 Butyrylcarnitine
C5 Isovalerylcarnitine
C6 Hexanoylcarnitine
C8 Octanoylcarnitine
C10 Decanoylcarnitine
C12 Dodecanoylcarnitine
CADTH Canadian Agency for Drugs and Technologies in Health
CAH Congenital adrenal hyperplasia
CDC Centers for Disease Control and Prevention
CID Collision-induced dissociation
CPTII Carnitine palmitoyltransferase II deficiency
DNA Deoxyribonucleic acid
EAIC Equivalent annual incremental cost
ESI Electrospray ionization
FTE Full-time equivalent
FAH Fumarylacetoacetate hydrolase
GAI Glutaric acidemia type I
GAIi Glutaric acidemia type II
HMG 3-hydroxy-3-methylglutaryl-CoA lyase deficiency
HMO Health maintenance organization
HPA Hyperphenylalaninemia
<table>
<thead>
<tr>
<th>Acronym</th>
<th>Full Form</th>
</tr>
</thead>
<tbody>
<tr>
<td>HPLC</td>
<td>High-performance liquid chromatography</td>
</tr>
<tr>
<td>ICER</td>
<td>Incremental cost-effectiveness ratio</td>
</tr>
<tr>
<td>INSPQ</td>
<td>Institut national de santé publique du Québec</td>
</tr>
<tr>
<td>IVA</td>
<td>Isovaleric acidemia</td>
</tr>
<tr>
<td>LC</td>
<td>Liquid chromatography</td>
</tr>
<tr>
<td>LCHADD</td>
<td>Long-chain hydroxyacyl-CoA dehydrogenase deficiency</td>
</tr>
<tr>
<td>MADD</td>
<td>Multiple acyl-CoA dehydrogenase deficiency</td>
</tr>
<tr>
<td>MAS</td>
<td>Medical Advisory Secretariat</td>
</tr>
<tr>
<td>MCAD</td>
<td>Medium-chain acyl-CoA dehydrogenase</td>
</tr>
<tr>
<td>MCADD</td>
<td>Medium-chain acyl-CoA dehydrogenase deficiency</td>
</tr>
<tr>
<td>3MCC</td>
<td>3-methylcrotonyl-CoA carboxylase deficiency</td>
</tr>
<tr>
<td>MMA</td>
<td>Methylmalonic acidemia</td>
</tr>
<tr>
<td>MRM</td>
<td>Multiple-reaction monitoring</td>
</tr>
<tr>
<td>M/SCHADD</td>
<td>Medium/short-chain L-3-hydroxyacyl-CoA dehydrogenase deficiency</td>
</tr>
<tr>
<td>MS/MS</td>
<td>Tandem mass spectrometry</td>
</tr>
<tr>
<td>MSSS</td>
<td>Ministère de la Santé et des Services sociaux</td>
</tr>
<tr>
<td>MSUD</td>
<td>Maple syrup urine disease</td>
</tr>
<tr>
<td>m/z</td>
<td>Molecular mass/charge (ratio)</td>
</tr>
<tr>
<td>NHMRC</td>
<td>National Health and Medical Research Council</td>
</tr>
<tr>
<td>NTBC</td>
<td>[2-(2-nitro-4-trifluoromethylbenzoyl)-1,3-cyclohexanedione]</td>
</tr>
<tr>
<td>PA</td>
<td>Propionic acidemia</td>
</tr>
<tr>
<td>PAH</td>
<td>Phenylalanine hydroxylase</td>
</tr>
<tr>
<td>Phe</td>
<td>Phenylalanine</td>
</tr>
<tr>
<td>PKU</td>
<td>Phenylketonuria</td>
</tr>
<tr>
<td>QALY</td>
<td>Quality-adjusted life-year</td>
</tr>
<tr>
<td>SAC</td>
<td>Succinylacetone</td>
</tr>
<tr>
<td>SCADD</td>
<td>Short-chain acyl-CoA dehydrogenase deficiency</td>
</tr>
<tr>
<td>SOP</td>
<td>Standard operating procedure</td>
</tr>
<tr>
<td>SRM</td>
<td>Single-reaction monitoring</td>
</tr>
<tr>
<td>TFP</td>
<td>Trifunctional protein deficiency</td>
</tr>
<tr>
<td>TT1</td>
<td>Tyrosinemia type 1</td>
</tr>
<tr>
<td>VLCADD</td>
<td>Very-long-chain acyl-CoA dehydrogenase deficiency</td>
</tr>
</tbody>
</table>
Glossary

Alleles
Alternate forms of a given gene that differ in their nucleotide sequence.

Analytical validity
The ability of a test to measure the property or characteristic that it is designed to measure.

Antibiogram
A bacteriological test aiming to evaluate the sensitivity of a given bacterium to various antibiotics.

Clinical validity
The ability of a test to diagnose a disease or predict future disease.

Composite heterozygote
A genotype or, by extension, an individual with two different mutated alleles at the same locus.

Confidence interval (CI)
A numerical interval within which the true parameter (mean, proportion or rate) is likely to fall with a predetermined level of probability (e.g., 95%).

Control or comparison group
A group of subjects observed during a study that serves as a basis of comparison for evaluating the effects of an exposure or intervention.

Cost-effectiveness analysis
An economic evaluation method involving the comparison of different intervention modalities, in which the costs are expressed in monetary units and the efficacy indicators in nonmonetary (natural) units.

Cost-utility analysis
An economic evaluation method involving the comparison of different intervention modalities, in which the costs are expressed in monetary units and the efficacy indicators in quality-adjusted life-years (QALYs).

Cut-off value
A value determining the boundaries for classifying the results of a test into values considered as normal and abnormal.

Daughter or product ions
Neutral or (positively or negatively) charged fragments resulting from the fragmentation of parent or precursor ions inside the tandem mass spectrometer collision cell.

Efficacy
The efficacy of a drug or an intervention reflects the extent to which a favourable outcome is achieved under ideal circumstances.

Efficiency
The efficiency of a treatment or an intervention reflects the extent to which a favourable outcome is achieved with the available resources.

External validity
The external validity of a study is the extent to which its results can be generalized to populations other than the population recruited for the study.
**Founder effect**
A high frequency of one or more mutations in the descendents of a small group of common ancestors, as a result of geographic and/or ethnic isolation.

**Gene**
A physical and functional unit of heredity consisting of a sequence of nucleotides situated at a specific locus on a given chromosome and generally coding for a protein that has a specific function.

**Genotype**
An individual’s genetic makeup, as contrasted with his or her phenotype. By extension, the genetic constitution at one or more specific loci.

**Heterozygote**
A genotype or, by extension, an individual with two different alleles at a given locus, either two mutated alleles in the case of composite heterozygotes, or one mutated and one non mutated allele.

**Homozygote**
A genotype or, by extension, an individual with two identical alleles at a given locus.

**Incidence**
A measure of the number of new cases of a disease appearing during a given period of time in a given population.

**Incorporation bias**
An error due to the incorporation of the results of the index test into the diagnostic workup (reference test), the concordance between the results of the two tests leading to an overestimation of the performance criteria of the index test.

**Meta-analysis**
A statistical method used to combine the results from different studies in order to obtain a quantitative estimate of the effect of a given exposure or intervention on a given outcome.

**Mutation**
Any change occurring in the DNA sequence that can result in pathological manifestations.

**Negative predictive value**
The probability that individuals with negative test results do not have the disease or will not develop the disease.

**Parent or precursor ions**
Intact molecules from the initial mixture to be analyzed by tandem mass spectrometry, which, once ionized, are separated and quantified in the first mass spectrometer.

**Penetrance**
The percentage of individuals with a specific genotype in whom the phenotype associated with this genotype is expressed.

**Phenotype**
The outward manifestation of the genotype in the form of a morphological trait, clinical syndrome or physiological characteristics, such as qualitative or quantitative variations in the final expression product of a gene (protein or metabolite).

**Positive predictive value**
The probability that individuals with a positive test result have the disease or will develop the disease.
Prevalence
A measure of the number of cases of a disease existing at a given point in time within a given population, including both old and new cases.

Prospective study
A study design is prospective if a group of individuals is followed in order to detect the occurrence of a disease or another outcome of interest.

Retrospective study
A study design is retrospective if part of the data collected for the analysis, whether on exposures or outcomes of interest, concern events that occurred before the study’s initiation.

Selection bias
An error due to systematic differences in characteristics between individuals included in a study and those who were not.

Sensitivity
A performance criterion of a diagnostic or screening test that measures its ability to correctly identify individuals with a given disease (or risk factor or health problem). The sensitivity of a test is the proportion of individuals with the target condition who have a positive test result.

Sequencing
The determination of the linear order of the DNA components.

Specificity
A performance criterion of a diagnostic or screening test that measures its ability to correctly identify individuals without a given disease (or risk factor or health problem). The specificity of a test is the proportion of individuals free of the target condition who have a negative result.

Validity
The validity of a measurement instrument reflects its ability to measure what it is designed to measure, and the validity of a study is the extent to which it is free of bias.
## CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>MISSION</td>
<td>i</td>
</tr>
<tr>
<td>FOREWORD</td>
<td>iii</td>
</tr>
<tr>
<td>ACKNOWLEDGEMENTS</td>
<td>iv</td>
</tr>
<tr>
<td>SUMMARY</td>
<td>vi</td>
</tr>
<tr>
<td>ABBREVIATIONS AND ACRONYMANS</td>
<td>xi</td>
</tr>
<tr>
<td>GLOSSARY</td>
<td>xiii</td>
</tr>
<tr>
<td>1 INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>2 OBJECTIVES AND METHODS OF THE LITERATURE REVIEW</td>
<td>2</td>
</tr>
<tr>
<td>3 REVIEW OF THE DISEASES</td>
<td>4</td>
</tr>
<tr>
<td>4 TANDEM MASS SPECTROMETRY</td>
<td>8</td>
</tr>
<tr>
<td>5 MS/MS PERFORMANCE FOR NEONATAL SCREENING OF INBORN ERRORS OF METABOLISM</td>
<td>10</td>
</tr>
<tr>
<td>6 ECONOMIC ASPECTS</td>
<td>23</td>
</tr>
<tr>
<td>7 ETHICAL, PSYCHOSOCIAL AND ORGANIZATIONAL ISSUES ASSOCIATED WITH NEONATAL SCREENING</td>
<td>28</td>
</tr>
<tr>
<td>8 DISCUSSION</td>
<td>30</td>
</tr>
<tr>
<td>9 CONCLUSION</td>
<td>36</td>
</tr>
<tr>
<td>APPENDIX A Elegibility criteria for study selection</td>
<td>39</td>
</tr>
<tr>
<td>APPENDIX B Prevalence of PKU</td>
<td>41</td>
</tr>
<tr>
<td>APPENDIX C Prevalence of TT1</td>
<td>42</td>
</tr>
<tr>
<td>APPENDIX D Prevalence of MCADD and of the A985G mutation</td>
<td>43</td>
</tr>
</tbody>
</table>
1 Introduction

Inborn errors of metabolism are hereditary metabolic diseases caused by mutations in genes coding for enzymes involved in a number of major metabolic pathways. Most of these diseases affect amino acid, fatty acid or organic acid metabolism. The mode of inheritance of mutations is most often autosomal recessive. Although the incidence of each of these diseases is low (1:10,000-1:1,000,000), their combined incidence (1:3,500) is high. The burden of these diseases in terms of infant morbidity and mortality is therefore substantial [Marsden et al., 2006; Raghuveer et al., 2006; Millington, 2002].

Quebec’s neonatal blood screening program for genetic diseases targets all newborns in the province in order to detect three genetic diseases, including phenylketonuria (PKU), since 1969; tyrosinemia type 1 (TT1), since 1970; and congenital hypothyroidism, since 1974. The latter is not discussed in this report, since it is not amenable to screening with tandem mass spectrometry (MS/MS). PKU screening is based on phenylalanine quantification by fluorometry. TT1 screening requires two first tier tests, namely, a semiquantitative succinylacetone assay and a fluorometric tyrosine assay, as well as a second tier quantitative succinylacetone assay [Laflamme et al., 2006].

The use of MS/MS technology for analyzing amino acids and acylcarnitines in newborns dates back to the early 1990s [Banta-Wright and Steiner, 2004; Carpenter and Wiley, 2002; Millington et al., 1990]. MS/MS has the potential to screen for more than 30 inborn errors of metabolism in a single analytical step [Chace et al., 1996; 1995; 1993; Rashed et al., 1995]. Some of these diseases cannot be detected by other technologies. The test, performed on a dried blood spot obtained by a heel prick of the newborn a few days after birth, generally takes about two minutes [Clarke, 2002].

A number of European countries and American states, as well as certain Canadian provinces recently reevaluated their neonatal screening programs and considered the relevance of introducing MS/MS and of expanding the list of diseases screened for [Lukacs and Santer, 2006; Tran et al., 2006; Watson et al., 2006; Health Council of the Netherlands, 2005; Pandor et al., 2004; MAS, 2002; Pollitt et al., 1997; Seymour et al., 1997]. In several jurisdictions, this re-evaluation led to the adoption of the technology and to an increase in the number of inborn errors of metabolism included in neonatal screening programs, this number ranging from 2 to more than 29, depending on the location [Garg and Dasouki, 2006]. In Canada, decisions regarding neonatal screening are of provincial jurisdiction. Seven Canadian provinces use MS/MS to screen newborns for 3 to 28 inborn errors of metabolism or are on the verge of introducing this technology. However, few jurisdictions have based their decisions on a systematic and rigorous evaluation of the evidence [Pollitt, 2006; Wilcken, 2006]. It is in this context that the MSSS asked AETMIS to evaluate the relevance of introducing MS/MS in Quebec for neonatal screening of inborn errors of metabolism.
OBJECTIVES AND METHODS OF THE LITERATURE REVIEW

After a close examination of the systematic reviews and an overview of the available Quebec data, it was agreed that AETMIS would first examine the relevance of replacing current methods for PKU and TT1 neonatal screening with MS/MS and of introducing neonatal screening for medium-chain acyl-CoA dehydrogenase deficiency (MCADD). The expansion of neonatal screening to other diseases could possibly be the subject of a subsequent report. The current report, which is a summary of a technical report [Makni et al., 2007] prepared at the MSSS’s request, describes, first, the methodology used to review the evidence, then a review of the natural history, epidemiology and management of the three diseases of interest. This is followed by a description of the MS/MS technology and a critical analysis of the evidence on its performance, an analysis of the capital and operating costs, and a discussion of the ethical, psychosocial and organizational issues associated with MS/MS implementation. Lastly, the discussion and the conclusions that can be drawn from this review are presented.

This report is based on the usual approaches to health technology assessment and proceeds by means of a critical review of the evidence, for each of the questions being addressed, which is completed by a contextual analysis pertinent to the Quebec situation. It should be noted that it does not cover all of the criteria or issues relevant to the decision-making process regarding neonatal screening programs. Solving the complex ethical and legal problems raised by neonatal screening in general lies beyond the scope of this study and requires a wider societal debate.

2.1 Literature search strategies

Separate literature searches were performed for the different sections of this report: the performance of MS/MS for neonatal screening of inborn errors of metabolism, the review of each of the diseases of interest (PKU, TT1 and MCADD) and the economic analyses. Of the articles identified by these search strategies, those addressing ethical, psychosocial and organizational issues were used for the discussion of issues associated with MS/MS-based neonatal screening (Chapter 7). The main literature search was conducted between July and December 2005. Subsequently, monthly updates were instituted until August 20066. The main search covered three databases, namely, Embase, PubMed and Cochrane, while only the latter two were used for the monthly updates. In addition, a search of the gray literature was performed with various search engines. Internet sites relevant to neonatal screening and genetics were also explored, as were specialized health technology assessment databases.

The literature search strategies are described in the technical report [Makni et al., 2007], including the keyword combinations used and the filters applied to focus the review of the three diseases on the themes “screening”, “diagnosis”, “prognosis”, “incidence/prevalence”, “epidemiology” and “treatment”. With a few exceptions, all search strategies were limited to the literature published as of 2000. For the section on MS/MS performance and for the theme «screening» in the review of the three diseases, the literature published as of 1995 was examined, while for the themes «incidence/prevalence» for TT1 and MCADD, no lower cut-off date was imposed. Restrictions in terms of publication type were applied as well: only review articles, consensus statements and recommendations were selected for the themes «diagnosis»

6. We did verify, however, whether any key articles were published between September 2006 and January 2007.
7. In the literature, the terms incidence and prevalence are used interchangeably for inborn errors of metabolism.
and «prognosis» for all three diseases and for the themes «incidence/prevalence», «epidemiology» and «treatment» for PKU. Lastly, for the themes “MS/MS performance” and «incidence/prevalence» and «treatment» of the three diseases, the reference list of selected key articles were scanned in order to identify other relevant articles.

2.2 Article selection and data extraction
Relevant articles were selected by one reviewer according to the eligibility criteria listed in Appendix A. Data extraction was carried out systematically using preestablished extraction forms for studies on MS/MS performance and tables for studies on disease incidence/prevalence and treatment. The extraction was also carried out by one reviewer, who consulted a scientific advisor to clarify interpretation problems, and contacted the authors of the article when this was deemed necessary. In many cases, we were unable to clarify the figures presented, despite several attempts to contact the authors. As a result, one study was excluded altogether [Yoon et al., 2005], whereas for others, several scenarios were considered in the calculation of performance criteria.

2.3 Study quality assessment
For the studies on MS/MS performance, a list of quality criteria was developed on the basis of the QUADAS tool [Whiting et al., 2006; 2003] and of the National Health and Medical Research Council recommendations [NHMRC, 2000]. No grading was used to appreciate the quality of the data, but each of the selected criteria was discussed thoroughly in order to produce a detailed assessment of the quality of the literature reviewed.

---

8. These figures pertain, for instance, to the number of subjects lost to follow-up or in whom the diagnosis has not been confirmed.
The clinical and genetic aspects of the three diseases of interest, namely PKU, TT1 and MCADD, are briefly presented here. They are autosomal recessive diseases. The epidemiological data for Quebec and Canada are provided when available. Recent therapeutic efficacy data are also discussed, since the contribution of neonatal screening to the prognosis depends to a large extent on the efficacy of early management.

### 3.1 Phenylketonuria

Phenylketonuria (PKU) is a hereditary disease caused by a deficiency in the phenylalanine hydroxylase enzyme, which leads to an increased blood phenylalanine level, or hyperphenylalaninemia. Untreated patients develop mental retardation and serious neurological problems. The gene involved in this disease is *PAH*, the gene coding for the phenylalanine hydroxylase enzyme, in which at least 524 mutations have been identified. Among the conditions caused by these mutations, two clinical forms are generally defined, namely PKU and non-PKU hyperphenylalaninemia. When a screening test result is positive, confirmation tests must be undertaken in order to determine the exact cause of the hyperphenylalaninemia in the newborn. Indeed, while a low-phenylalanine diet should be instituted as soon as possible in PKU, a lifetime diet is not required in non-PKU hyperphenylalaninemia. According to the latest Quebec data, non-PKU hyperphenylalaninemia accounts for up to 50% of all the cases of hyperphenylalaninemia [Laflamme et al., 2006].

Patients who are identified early and in whom the serum phenylalanine level normalizes with treatment do not experience any neurological damage and have a normal intelligence quotient. Thus, the benefits of early treatment in PKU are well established. It is now recommended that patients continue this treatment for life [Feillet, 2006; Koch et al., 2002]. A follow-up by specialists is necessary in order to monitor the serum phenylalanine levels and provide the necessary support and information [Abadie et al., 2005; Camfield et al., 2004; Donlon et al., 2004; Hanley, 2004]. Women of child-bearing age require close attention, since maternal hyperphenylalaninemia, whether resulting from PKU or non-PKU hyperphenylalaninemia, can adversely affect fetal development.

The incidence of PKU is not the same throughout the world and is apparently higher in European and Chinese populations. As for the situation in Canada, one study estimated the incidence of PKU to be 4.5 cases per 100,000 births. Estimates vary from province to province however (Appendix B). In Quebec, the prevalence is 4 cases per 100,000 births, which corresponds to three affected newborns annually [Laflamme et al., 2006].

Although PKU is one of the better known inborn errors of metabolism, there are still a number of questions regarding its treatment and prognosis. These questions concern, among other things, the long-term neurological development [Burgard et al., 2000], the continuation of the diet during adulthood [Hanley, 2004], the monitoring of affected

---

9. An autosomal recessive inheritance pattern means that an affected person has two mutated genes, one received from the mother and the other from the father. The parents are usually not affected but healthy carriers, each having one mutated and one normal gene.

10. The distinction is based on different serum phenylalanine levels and on the tolerance to dietary phenylalanine. It should be noted that the nomenclature used in the literature may vary. Thus, PKU may be subdivided into classical PKU and atypical PKU. Non-PKU hyperphenylalaninemia may also be referred to as “benign”.
pregnant women [Kalter, 2003], the use of dietary protein supplements [Rutherford and Poustie, 2005], the improvement in the composition and quality of dietary products [Feillet, 2006; Macdonald et al., 2004; van Spronsen et al., 2001a; 2001b], the adherence to the diet [Mackner et al., 2001], the pathophysiology of PKU [NIH Consensus Development Panel, 2001] and the efficacy of alternative treatments [Feillet, 2006; Spaapen and Rubio-Gozalbo, 2003; NIH Consensus Development Panel, 2001].

Systematic reviews on the efficacy of PKU treatment reveal that the available studies are of sub-optimal quality, but the relevance of neonatal screening and early treatment is not brought into question [Rutherford and Poustie, 2005; Macdonald et al., 2004; van Spronsen et al., 2001a; 2001b; Poustie and Rutherford, 2000a; 2000b]. PKU screening, which has been in place in many countries for nearly 40 years, is often considered a model with regard to the goals and results of neonatal screening, [Donlon et al., 2004].

3.2 Tyrosinemia type 1

Tyrosinemia type 1 (TT1) is a metabolic disease due to a deficiency in fumarylacetoacetate hydrolase (FAH), the final enzyme in the tyrosine catabolic pathway. This deficiency leads to an accumulation of tyrosine metabolites, one of which is succinylacetone. Such an accumulation results in renal and hepatic toxic effects and in neurological crises. Some 44 mutations have been identified in the gene coding for FAH, but one mutation at intron 12 (IVS12+5g → a) accounts for most of the cases of TT1 in the French-Canadian population and for about a third of the cases worldwide.

Two clinical forms of TT1 have been defined on the basis of disease severity and age at diagnosis. The acute form, which accounts for about 75% of cases, usually manifests during the first weeks of the infant’s life with signs of severe hepatic failure. If left untreated, most of these children die within the first year of life [Ashorn et al., 2006; Scott, 2006; Mitchell et al., 2001]. The chronic form usually manifests after the age of 6 months. It is characterized by hepatic and renal dysfunction and insufficient weight gain. The risk of developing hepatocarcinoma can be as high as 37% in untreated children [Scott, 2006; Mitchell et al., 2001]. In both clinical forms, children can present with severe neurological crises, which is a major cause of mortality and morbidity.

With a restricted diet, the 2-year survival rate in affected children is 29% when the first symptoms appear before the age of 2 months, 74% when they appear between the ages of 2 and 6 months, and 96% when they appear after the age of 6 months (chronic form). However, the 10-year survival rate drops to 30% in the group diagnosed between the ages of 2 to 6 months and to about 60% in the group diagnosed after the age of 6 months [van Spronsen et al., 1994].

The worldwide prevalence of TT1 is estimated at about 1 case per 100,000 to 120,000 births. The prevalence of the disease is especially high in Scandinavia and Quebec [Mitchell et al., 2001]. As indicated in Appendix C, the Quebec prevalence is estimated at 6 per 100,000, which corresponds to four affected newborns a year, while it is 54 per 100,000 in the Saguenay-Lac-Saint-Jean region because of a founder effect11 [De Braekeleer and Larochelle, 1990].

In the 1970s, TT1 treatment consisted in initiating a very strict diet low in tyrosine, phenylalanine and methionine as soon as the disease was diagnosed. In the decade that followed, liver transplantation emerged as the only effective therapeutic option.

---

11. The prevalence calculation for the Saguenay-Lac-Saint-Jean region is based on data from 1982 to 1986, which showed the rate of heterozygous carriers of the founder mutation to be 1 in 21 [De Braekeleer and Larochelle, 1990].
[Ashorn et al., 2006; Scott, 2006]. Presently, the treatment consists of a NTBC\textsuperscript{12}-based pharmacologic therapy along with a protein-restricted diet, liver transplantation being performed only if the NTBC therapy fails. A large, multicentric clinical study involving more than 300 patients treated with NTBC around the world is underway. The preliminary results indicate an improvement in the clinical picture in 90\% of patients with the acute form as compared to 75\% of patients with the chronic form [Holme and Lindstedt, 2000]. An improvement in the patients’ hepatic function and metabolic profile and a decrease in the number of hepatocarcinomas were observed [Holme and Lindstedt, 2000; 1998]. A preliminary analysis of the Quebec data, collected as part of this multicentric clinical trial, seems to indicate a statistically significant decrease in the number of liver transplants and of deaths\textsuperscript{13} [Quebec NTBC Study Group, 2005]. These results emphasize the importance of early diagnosis and management. However, the long-term consequences of the treatment are still largely unknown, given the limited time NTBC has been in use. Although universal neonatal TT1 screening in Quebec was instituted before an effective treatment was available, the data accumulated over the past few years, even though still preliminary, seem to show that there is a real benefit in screening for this disease.

3.3 Medium-chain acyl-CoA dehydrogenase deficiency

MCAD (medium-chain acyl-CoA dehydrogenase) deficiency, or MCADD, is caused by a defect in the MCAD enzyme, which plays an important role in the production of energy from fatty acids during fasting or following a metabolic stress [Grosse et al., 2006b; Goddard, 2004; Olpin, 2004; Roe and Ding, 2001]. MCAD deficiency is characterized by an accumulation of acylcarnitines\textsuperscript{14} in the plasma due to incomplete fatty acid catabolism. Thus far, more than 30 different mutations have been identified in the gene coding for this enzyme. The most common mutation (A985G) is found in the homozygous state, that is, on both alleles, in more than 80\% of the patients of European origin diagnosed on the basis of clinical symptoms. Fasting, exercise or intercurrent infectious diseases can cause acute symptoms with hypoglycemia and metabolic acidosis. Severe episodes of metabolic decompensation can progress to acute encephalopathy. The mortality rate associated with these episodes is high (up to 25\%), and neurological sequelae are relatively frequent [Grosse et al., 2006b; Goddard, 2004; Wilcken et al., 2002; Iafolla et al., 1994]. The disease sometimes manifests as an unexplained sudden infant death. The less severe forms present with hypotonia, lethargy and vomiting. Most symptomatic individuals develop signs of the disease between the ages of 3 months and 3 years, but according to some authors, one third to one fourth of individuals with MCADD remain asymptomatic throughout their lives [Grosse et al., 2006b; Goddard, 2004]. In fact, several studies suggest that the disease is underdiagnosed in populations where no neonatal screening program is in place [Hoffmann et al., 2004; Wilcken et al., 2003; Carpenter et al., 2001; Pourfarzam et al., 2001]. The diagnosis, which is not standardized, is based on several biochemical tests and on the expertise of metabolic

\textsuperscript{12} NTBC [2-(2-nitro-4-trifluoromethylbenzoyl)-1,3-cyclohexanedione], also called “nitisinone” (Orfadin*), belongs to a class of compounds used as herbicides in the 1980s.

\textsuperscript{13} In fact, transplants were necessary in 19 of the 27 untreated patients (70.3\%) compared to 5 of the 52 treated patients (9.6\%). The number of deaths associated with the disease also decreased significantly. There were 10 deaths in the non-NTBC treatment group as opposed to two in the treatment group ($p < 0.0001$), both of which were due to post-transplant complications in patients whose treatment had not been initiated early [Quebec NTBC Study Group, 2005].

\textsuperscript{14} Acylcarnitines are identified by the length of the acyl-CoA chain. For example, C8 stands for octanoylcarnitine, C10 for decanoylcarnitine.
disease specialists. Treatment, which is aimed at minimizing the risks associated with fasting, involves frequent food intake, especially during the first few months of life. The management of the symptomatic forms through simple dietary measures prevents the recurrence of acute episodes to a large extent. However, given the large variability in disease spectrum, it is difficult to draw conclusions about the benefits of screening and of early management for all patients undergoing screening on the basis of the follow-up of clinically diagnosed patients.

Prevalence data, estimated from MS/MS-based neonatal screening programs, are presented in Appendix D. The incidence is high, especially in European populations. From the international literature, the prevalence in Canada has been estimated to be 6:100,000 [Tran et al., 2006]. Also provided in Appendix D are data on the proportion of A985G mutation carriers in the population. One Manitoban study reports a carrier rate of 1:154 [Thompson et al., 1995], while in Quebec, the results of an ongoing study involving more than 6000 newborns points to a carrier rate of approximately 1:7215.

With the implementation of MS/MS-based MCADD screening programs16, data on the prognosis of patients screened and managed before the onset of symptoms are starting to accumulate. Tran and colleagues [2006] reviewed clinical data drawn from nine prospective cohort studies on MS/MS-based screening and compared them to those from two retrospective studies involving series of clinically diagnosed patients. Our literature update revealed one additional study of each type [Derks, 2006; Nennstiel-Ratzel et al., 2005]. The proportion of patients who experience metabolic crises at a young age and the number of deaths are lower in newborns identified by neonatal screening than in children diagnosed clinically17. The variability in clinical manifestations makes it more difficult to compare the prognosis with and without screening and early management. Molecular analyses show that the proportion of homozygous children for the common mutation A985G is notably different in cohorts of clinically diagnosed children (approximately 80%) than in those identified by screening (about 40% to 70%). Even though the genotype/phenotype correlation is still very poorly understood and the clinical expression is not determined solely by the genotype, these differences tend to confirm that a direct comparison of outcomes between such cohorts is prone to selection bias.

With the current evidence, and despite its limitations, the balance between the pros and cons is in favour of neonatal screening. Indeed, the benefits of an early treatment for severely affected patients are so substantial that they seem to outweigh the uncertainty surrounding the benefits for those less severely affected. In addition, a number of modelling exercises, which admittedly have their limitations, provide an additional argument, namely that of the cost-effectiveness of screening.

---

15. Dr. F. Rousseau, Full Professor, Department of Medical Biology, Faculty of Medicine, Laval University, Quebec, personal communication, November 7, 2006.

16 MS/MS technology is the only means of screening for MCADD.

17. Tran and colleagues’ results [2006] tend to indicate that acute symptoms occur more frequently in patients diagnosed without screening (76%; 95% CI: 64-85) than in children identified by neonatal screening (4%; 95% CI: 1-9). This difference is statistically significant. Similarly, less than 2% of deaths is to be expected after neonatal screening, while a proportion of 16% of deaths is recorded in studies based on a clinical diagnosis. It should be noted that the number of children with MCADD was 8 and 62, respectively, in the two studies not involving neonatal screening, and ranged from 2 to 41 in the studies involving MS/MS-based neonatal screening.
Mass spectrometry is essentially used to qualitatively and quantitatively analyze several types of complex mixtures of biological metabolites. The identification of these metabolites is based on the measurement of the mass of the molecules and of their ionized fragments, which are separated and quantified on the basis of their molecular mass/charge ratio (m/z) [Banta-Wright and Steiner, 2004; Cheillan et al., 2004; Clarke, 2002]. A tandem mass spectrometer comprises two mass spectrometers separated by a collision cell, in which the target metabolites are fragmented. The particularity of this device is that the first spectrometer quantifies the intact molecules from the initial mixture, while the second spectrometer quantifies their fragments. The information obtained from both mass spectrometers is then matched, allowing for the identification of the target metabolites [Cheillan et al., 2004; Rinaldo et al., 2004; Chace et al., 2003; Dooley, 2003; Carpenter and Wiley, 2002; Clarke, 2002, Millington, 2002]. An illustration and a brief description of the MS/MS device are provided in Appendix E. The technical report contains a more detailed description of its operating modes, enabling a more thorough understanding of the limitations of this technology [Makni et al., 2007]. The results of the MS/MS analyses are presented in a graph called a «mass spectrum», where the abscissa represents the different m/z ratios, and the ordinate represents the quantity of ions. The complexity of the MS/MS technology relates to the diversity of operating programs and analytical modalities. This technology can, for example, be used to perform a full scan of all metabolic profiles associated with a family of inborn errors of metabolism or to selectively screen for specific diseases (see Appendix E) [Cheillan et al., 2004; McCandless, 2004; Rinaldo et al., 2004; Dooley, 2003; Carpenter and Wiley, 2002; Millington, 2002; Elgstoen et al., 2001].

The full scan protocol which is most often used in neonatal screening of inborn errors of metabolism is the common analysis of amino acid and acylcarnitine profiles. This protocol enables one to simultaneously screen for more than 30 inborn errors of amino acid, fatty acid and organic acid metabolism [Garg and Dasouki, 2006; Cheillan et al., 2004; Carpenter and Wiley, 2002; Chace et al., 2002; Rashed et al., 1995]. In fact, this protocol uses a high-speed alternation between three scanning modes. Appendix F summarizes in tabular form all of the required steps from sample preparation, which involves the extraction and derivatization of the target metabolites, to the production of results.

The selective analysis of inborn errors of metabolism is based on the preferential detection of ions with a specific m/z ratio. The analytical mode referred to as «Single-Reaction Monitoring» (SRM) is used to quantify a specific metabolite, but it is also possible to simultaneously perform several SRMs in order to quantify more than one metabolite of interest [Chace et al., 2005; Dooley, 2003]. This analytical mode is used, for example, in the common analytical protocol for amino acids and acylcarnitines in order to target certain inborn errors of metabolism, while avoiding the detection of others for which the natural course and/or treatment is less well known. In addition, the SRM mode is used to analyze metabolites that are not detected by the common protocol for amino acids and acylcarnitines. Description of such an SRM protocol for succinylacetone quantification, which is necessary for TT1 screening, was recently published [Allard et al., 2004]. This protocol requires an additional extraction step involving the use of residual dried blood spots that are left over from the amino acid and acylcarnitine extraction, as well as a separate MS/MS analysis. Research
is underway in order to combine into a single step the simultaneous quantification of succinylacetone and of the metabolites of interest usually detected by the common amino acid and acylcarnitine protocol [Allard, 2005]. However, these are preliminary investigations and this approach requires further rigorous validation.

The fact that this technology is automated makes it suitable for a high throughput, of up to 600 samples per 24-hour period [Cheillan et al., 2004; Chace et al., 2003; Fearing and Levy, 2003; Carpenter and Wiley, 2002; Clarke, 2002]. Furthermore, very small quantities of metabolites can be detected, separated and identified. Aside from the recognised advantages of the MS/MS technology, a number of drawbacks have also been reported [Cheillan et al., 2004; Chace et al., 2003; Fearing and Levy, 2003; Carpenter and Wiley, 2002; Clarke, 2002]. These primarily concern the technical requirements, which are rather stringent, with the need to take precautions at all steps, from blood sampling to results interpretation. Many authors mention the instrument’s fragility and the need for each laboratory\textsuperscript{18} to have a backup instrument so as to avoid screening interruption during machine downtime. Certain limitations are related to the specific use of MS/MS in neonatal screening, either because of the nature of the metabolite or because of the timing of sampling. Lastly, a number of interpretation problems arise in terms of differential diagnosis between the targeted metabolites and contaminants or between different diseases [Chace et al., 2005]. The advantages and disadvantages of the MS/MS technology are explained in greater detail in Appendix G.

The literature indicates that many technical features, related to the different analytical steps, can affect MS/MS performance [CDC, 2001]. Yet, recent publications report a rather wide variability in technological choices from one neonatal screening centre to another, which further complicates the comparison and interpretation of MS/MS performance data.

The technology is constantly evolving in terms of the analytical processes. Among the foreseen technological changes that could substantially alter MS/MS performance is the incorporation of the succinylacetone assay into the protocol presently used to analyze tyrosine [Allard, 2005] and the use of underivatized analytical protocols [Garg and Dasouki, 2006; Schulze et al., 2003b; Trinh et al., 2003; Qu et al., 2002]. The performance of these new approaches will need to be rigorously evaluated prior to implementation. In addition, the evolving nature of the technology underscores the importance of carefully considering technological options prior to implementation, of performing analytical validation following any change to the protocol, and of establishing ongoing quality assurance mechanisms.

\textsuperscript{18} The annual number of analyses per laboratory recommended by Pandor and colleagues [2004] is 50,000 to 60,000. However, the instrument has a higher analytical capacity [CDC, 2001]. Thus, one laboratory should be sufficient for the annual number of births in Quebec, which is approximately of 75,000. At the time this report was in press, a government document published in January 2007 indicated that the annual number of births may have reached 82,500 in 2006, a substantial increase over previous figures [MFACF, 2007].
Our review of the evidence on MS/MS performance for neonatal screening is mainly based on an analysis of three recent systematic reviews [Pandor et al., 2006a; 2004; Tran et al., 2006] and of more recent primary studies. Special attention was given to studies reporting on the selective performance of MS/MS for the three diseases of interest.

5.1 Systematic reviews

Two British studies were published by the same authors in 2004 and 2006 [Pandor et al., 2006a; 2004], and a third was published in 2006 by the Canadian Agency for Drugs and Technologies in Health (CADTH) [Tran et al., 2006]. The main results and the recommendations of these systematic reviews are summarized below. Further details and a description of the methods, the literature covered and the limitations of these studies are included in the technical report [Makni et al., 2007].

The primary objective of the review of Pandor and colleagues [2004] was to update two British reviews published in 1997 [Pollitt et al., 1997; Seymour et al., 1997] and to evaluate the clinical efficacy and the cost-effectiveness of MS/MS-based neonatal screening of inborn errors of metabolism [Pandor et al., 2004]. The authors concluded that the current data derive mainly from observational studies conducted in the context of large-scale neonatal screening programs in Germany, Australia and the United States. They considered MS/MS to be rapid as well as highly sensitive (90-100%) and specific (99-100%) for the detection of groups of inborn errors of amino acid and acylcarnitine metabolism, but pointed out that there are not enough data to extend these conclusions to the neonatal screening of individual inborn errors of metabolism, except for MCADD (100% sensitivity and specificity). Performing a quantitative meta-analysis was considered but rejected because of the heterogeneity between studies. The major limitations of the literature included: 1) the difficulty in detecting and differentiating between certain metabolites and certain diseases, which can lead to false positives; 2) the paucity of studies with a sufficiently long follow-up to allow for an accurate determination of the proportion of false negatives; and 3) the difficulty in comparing studies because of differences in the age of sampling, the choice of metabolites or metabolite ratios, the cut-off values used, and the diagnostic confirmatory tests.

Given the performance data and the results of their economic analysis (see Chapter 6), Pandor and colleagues [2004] concluded that «The evidence appears to support the introduction of tandem MS into a UK neonatal screening programme for PKU and MCAD deficiency combined». They also identified key research areas requiring attention prior to expanding the British neonatal screening program to other diseases. Essentially, they underscored: 1) the need to improve the state of knowledge on MS/MS performance for neonatal screening of individual inborn errors of metabolism and on the natural course and treatment of poorly understood diseases; and 2) the importance of instituting long-term patient follow-up in order to determine as accurately as possible the proportions of false negatives.

19. Other reports were examined but were not used as a basis for our literature analysis either because they were updated or because they were not based on a systematic review of the evidence. These reports are, however, described in an appendix in the technical report [Makni et al., 2007].
In 2006, Pandor and colleagues carried out an update of their review, which they limited to studies reporting on MS/MS performance for the selective neonatal screening of PKU and MCADD [Pandor et al., 2006a]. They concluded that recent evidence is prone to the same limitations as previous studies and that it confirms the high sensitivity and specificity of MS/MS-based neonatal screening of PKU and MCADD. In addition, they stressed the importance of using the phenylalanine/tyrosine ratio in combination with phenylalanine quantification, as this results in a higher positive predictive value being obtained with MS/MS than with conventional methods for neonatal PKU screening.

The aim of the review of Tran and colleagues [2006] was to evaluate the potential application of MS/MS for MCADD screening in the Canadian context in light of the clinical, financial, ethical and psychosocial issues. They evaluated the quality of the studies using the QUADAS tool [Whiting et al., 2006; 2003] and pooled the data extracted from each study on sensitivity, specificity, positive predictive value and negative predictive value by calculating a weighted mean with a 95% confidence interval for each performance criterion, regardless of the heterogeneity between studies.

The authors concluded that most of the reviewed studies are of suboptimal quality, especially with regard to the external validity, because of the restriction of the study populations to specific ethnic groups and because of the lack of information on subject selection criteria in most of the studies. They also criticized the lack of long-term patient follow-up, which particularly affects the estimation of the false negative rate, and the lack of information on the diagnostic confirmatory protocols. In addition, few studies included a sufficiently detailed description of the MS/MS analytical protocol, and none of them provided data on uninterpretable test results or on subjects lost to follow-up.

As regards the performance criteria, the authors stated that the studies’ results indicate that MS/MS has maximum sensitivity and negative predictive value (100%) for neonatal MCADD screening, a specificity of 99.98% to 100% (weighted mean of 99.99%), and a positive predictive value of 19% to 100%, with a weighted mean of 51% (95% CI: 11-91). However, they explained that estimation of these criteria is based on the assumption of a null false negative rate, since the studies do not have a long enough follow-up period to estimate this rate more rigorously. Tran and colleagues [2006] therefore recommended the introduction of MS/MS-based neonatal screening for MCADD. However, they stressed the importance of developing a pan-Canadian consensus on the diagnostic workup of MCADD.

5.2 Primary studies

The search strategies for the themes «MS/MS performance» and «MS/MS-based screening for inborn errors of metabolism» yielded 306 references published before 2000 and 453 references published thereafter (Appendix H). However, upon reading the articles, it was found that only 13 met the selection criteria for this section of the report [Feuchtbaum et al., 2006b; Frazier et al., 2006; Sander et al., 2006; Schulze et al., 2003a; Wilcken et al., 2003; Ceglarek et al., 2002; Shigematsu et al., 2002; Andresen et al., 2001; Carpenter et al., 2001; Pourfarzam et al., 2001; Zytkovicz et al., 2001; Chace et al., 1998; 1997]. Two studies appear to have been funded by the private sector (Neo Gen Screening) [Chace et al., 1998; 1997]. A table presenting the general characteristics of these studies, including the country, the study period, the neonatal screening program, the origin of the population, and the study design, is provided in Appendix I. A detailed description of the selected studies and their limitations is provided in the technical report [Makni et al., 2007].

Of the selected studies, six investigated MS/MS-based neonatal screening for groups of inborn errors of metabolism [Feuchtbaum et al., 2006b; Frazier et al., 2006; Schulze
et al., 2003a; Wilcken et al., 2003; Shigematsu et al., 2002; Zytkovicz et al., 2001], while the other seven examined the use of this technology for the selective neonatal screening of PKU [Ceglarek et al., 2002; Chace et al., 1998], TT1 [Sander et al., 2006] or MCADD [Andresen et al., 2001; Carpenter et al., 2001; Pourfarzam et al., 2001; Chace et al., 1997]. It was however possible to extract MS/MS performance data for selective PKU, TT1 or MCADD screening from some of the studies that provided results for groups of inborn errors of metabolism. The results of the studies on groups of inborn errors of metabolism are discussed very briefly below, whereas the data extracted from these references on PKU, TT1 and MCADD are discussed along with the results of the studies that investigated each of these diseases specifically.

5.2.1 Study quality

In terms of study quality, a number of reservations are noteworthy, namely with respect to the study designs, the population selection procedures, the sampling and sample transport conditions, the diagnostic workup, the analytical methods, and the reporting of results. Most of the available evidence derives from prospective cohort studies carried out in the context of neonatal screening programs [Feuchtbau et al., 2006b; Frazier et al., 2006; Sander et al., 2006; Schulze et al., 2003a; Wilcken et al., 2003; Shigematsu et al., 2002; Andresen et al., 2001; Carpenter et al., 2001; Zytkovicz et al., 2001; Chace et al., 1997]. It is indeed difficult to conduct clinical trials or prospective controlled studies with an appropriate control group in the context of neonatal screening of inborn errors of metabolism because of the large sample size needed, given the rarity of each disease, and the length of follow-up required [Shortland, 2004; Wilcken et al., 2003]. A few retrospective studies have compared MS/MS performance to that of fluorometry, the method currently used for neonatal PKU screening in Quebec [Ceglarek et al., 2002; Chace et al., 1998].

The subject selection methods, the eligibility criteria used, the characteristics of the newborns in terms of ethnicity, gestational age, weight and health condition at birth, and the number of subjects lost to follow-up have seldom been described. These shortcomings, related to the study populations, limit the generalizability of the results. Apart from differences in disease prevalence between regions, the factors potentially most important for MS/MS are the percentage of newborns who have a low birthweight, who are premature, who have been treated with certain antibiotics or who have received transfusions, vitamins or parenteral hyperalimentation [Marsden et al., 2006; Chace et al., 2005; Shigematsu et al., 2002; Zytkovicz et al., 2001].

As for the sampling conditions, the age at sampling varies from study to study, and the duration of feeding at the time of specimen collection is never indicated. Yet, it is important to consider these factors, since metabolite concentrations change with time and with dietary intake during the neonatal period [Marsden et al., 2006; Chace et al., 2005; Zytkovicz et al., 2001].

Like Pandor and colleagues [2004], we noted wide variations in the choice of metabolic markers, cut-off values, protocols for classifying MS/MS results, and confirmatory tests for a given disease. These differences result in a substantial heterogeneity in the literature and render the comparison of results between studies difficult. The variations in classification protocols and confirmatory testing are related to the nature of this group of metabolic diseases, whose diagnoses often necessitate a series of tests, rather than a single test, and are generally based on a metabolic disease specialist’s interpretation
of all test results and of the patient’s clinical condition. A systematic comparison of the results of the index test with a gold standard, as is conventionally required in diagnostic test evaluations, is therefore difficult because of the number of inborn errors of metabolism, and also because of the nature of the diagnostic process.

In addition, the authors of only one study stated that interpretation of the MS/MS results was blinded to the fluorometry results [Chace et al., 1998]. Consequently, a review bias cannot be ruled out for the other studies, whether for the interpretation of MS/MS results in the retrospective studies or for the diagnostic classification in the prospective studies. Furthermore, in one study on MCADD [Carpenter and Wiley, 2002], the MS/MS results may have been included in the diagnostic confirmation protocol, which leads to an incorporation bias and an overestimation of the performance criteria.

In all of the studies, the confirmatory tests were performed only for the patients who had a positive MS/MS result, and the exclusion of false negatives was never based on rigorous methods applied to all, or at least to a random sample, of the newborns in the study. The sensitivity and negative predictive value of the MS/MS test may thus have been overestimated, which represents by far the main limitation of the reviewed studies.

5.2.2 Results of the selected studies

5.2.2.1 MS/MS performance for the neonatal screening of groups of inborn errors of metabolism

These studies are discussed in detail in the technical report [Makni et al., 2007], whereas only the highlights are presented here. Overall, the results of the different studies concur that MS/MS has a high specificity (99.43–99.99%) and a high negative predictive value (99.99–100%) (Appendix J). In contrast, the sensitivity (91.66–100%) and the positive predictive value (2.02–59.76%) vary widely from study to study, with recall rates ranging from 0.03% to 0.58% [Feuchtbaum et al., 2006b; Frazier et al., 2006; Schulze et al., 2003a; Wilcken et al., 2003; Shigematsu et al., 2002; Zytkovicz et al., 2001].

It should be noted that the number of inborn errors of metabolism considered in the studies varies and that the reporting of results is sometimes imprecise with regard to the classification of MS/MS results or the number of subjects lost to follow-up or for whom there was no formal diagnostic confirmation. Indeed, some authors used one series of cut-off values [Feuchtbaum et al., 2006b; Frazier et al., 2006; Schulze et al., 2003a; Wilcken et al., 2003; Shigematsu et al., 2002; Zytkovicz et al., 2001], while others used two cut-off values (one borderline and one diagnostic) [Frazier et al., 2006] or incorporated into the diagnostic confirmation protocol a medical biochemist’s evaluation...

---

20. The diagnostic approach generally involves a repeat MS/MS analysis of the dried blood sample [Frazier et al., 2006; Schulze et al., 2003a; Zytkovicz et al., 2001], a plasma amino acid and acylcarnitine profile [Frazier et al., 2006; Schulze et al., 2003a; Wilcken et al., 2003; Zytkovicz et al., 2001], a urinary assay of organic acids [Frazier et al., 2006; Schulze et al., 2003a; Wilcken et al., 2003; Zytkovicz et al., 2001] and/or amino acids [Wilcken et al., 2003], a DNA analysis to identify mutations, especially for MCADD [Frazier et al., 2006; Schulze et al., 2003a; Wilcken et al., 2003; Zytkovicz et al., 2001], and an evaluation of the deficient enzyme’s activity [Frazier et al., 2006; Schulze et al., 2003a; Wilcken et al., 2003]. For some targeted inborn errors of metabolism, Schulze and colleagues [2003a] also took into consideration the clinical course of the patients who had a positive MS/MS test result. The length of this clinical follow-up varied from 0.1 to 38 months, with a mean of 13.5 months, and was complemented by a diagnostic re-evaluation after 12 months of follow-up.

21. However, Schulze and colleagues [2003a] indicated that, up until their study was published, they had sent monthly questionnaires to all the German pediatric hospitals and metabolic centres in order to check if a case of inborn error of metabolism had been missed by the neonatal screening program during their study. Similarly, Feuchtbaum and colleagues [2006b] attempted to check their false-negative rates by collaborating with the California coroner, who identified all the cases of newborns who had died of unknown causes during their study period. The authors indicated that none of the 16 cases identified by the coroner was diagnosed with an inborn error of metabolism. None of the other studies reported such efforts to detect potential false negatives.
of the metabolic profile [Feuchtbauern et al., 2006b]. For Zytkovicz and colleagues’ study [2001], we distinguished between two groups of inborn errors of metabolism, based on the number of newborns tested22, so as not to limit the performance criteria calculation to the positive predictive value (Appendix J, Table J-1). In addition, many authors adjusted their cut-off values during the course of their study after having identified false negatives. Such false negative results have in fact been reported for infants with hyperphenylalaninemia (type not indicated) [Schulze et al., 2003a], TT1 [Frazier et al., 2006; Wilcken et al., 2003] as well as MCADD [Shigematsu et al., 2002].

5.2.2.2 MS/MS performance for the selective screening of PKU

MS/MS performance data on the selective neonatal screening of PKU or hyperphenylalaninemia were extracted from six studies. Two studies were specific to these inborn errors of metabolism and compared the MS/MS to the fluorometry results [Ceglarek et al., 2002; Chace et al., 1998], while the four others investigated groups of diseases [Frazier et al., 2006; Schulze et al., 2003a; Wilcken et al., 2003; Zytkovicz et al., 2001]. Table 1 presents the results of these six studies, indicating for each the number of true and false positives, of true and false negatives as well as the performance criteria extracted from the article or calculated according to the test (if applicable) and the metabolic marker(s) used. For Frazier and colleagues’ [2006] and Schulze and colleagues’ [2003a] studies, several calculation scenarios were considered because of the use of several cut-off values for the classification of the MS/MS results in the former and the uncertainty about the classification of a few subjects in the latter. It is also important to note that most authors count newborns with non-PKU hyperphenylalaninemia among the true positives. We followed this approach to ensure homogeneity in the treatment of the results of the different studies. Lastly, Chace and colleagues [1998] sought to evaluate the validity of the phenylalanine/tyrosine ratio in a population of infants less than 24 hours of age. Thus, this study involved a limited number of samples selected on the basis of age and of the fluorometry results.

In order to facilitate the interpretation of the results, we shall first discuss MS/MS performance for PKU screening by comparing the results of the six studies, then examine the comparisons between MS/MS and fluorometry results.

a) Performance results for MS/MS only

The number of samples analyzed varied from 203 to 362,000. Three studies [Ceglarek et al., 2002; Zytkovicz et al., 2001; Chace et al., 1998] provided results separately for the metabolic marker phenylalanine and the combination of this marker and the phenylalanine/tyrosine ratio; one study used phenylalanine quantification only [Wilcken et al., 2003]; and the remaining two only considered the combination of these two markers for the purpose of classifying MS/MS results [Frazier et al., 2006; Schulze et al., 2003a]. If we consider only the phenylalanine concentration as the metabolic marker, the results of the studies concur, showing a perfect sensitivity and negative predictive value (100%) and an excellent specificity (98.37-99.99%) for MS/MS [Wilcken et al., 2003; Ceglarek et al., 2002; Zytkovicz et al., 2001; Chace et al., 1998]. In contrast, the results diverge considerably with regard to the positive predictive value (2.34-86.36) and the recall rate (0.03-10.84). These differences are probably related, at least in part, to variations between studies in the prevalence of PKU, especially for Chace and colleagues’ study [1998], in which the prevalence was probably artificially high because of the study population selection methods. In addition, the use of different cut-off values may affect the test specificity, while even a tiny variation in test specificity causes considerable divergence in the positive predictive value, as pointed out by Tran and colleagues [2006].

---

22. The number of newborns tested was 257,000 for PKU, maple syrup urine disease and hypermethioninemia; 184,000 for MCADD; and 164,000 for TT1, deficiencies of ARG (arginase), ASS (arginosuccinate synthetase) and ASL (arginosuccinate lyase), and all inborn errors of acylcarnitine metabolism, with the exception of MCADD, respectively. Given that the results for MCADD pertain to a separate group, they were not included in the performance criteria calculation presented in Table J-1, but are presented in Section 5.2.2.4.
## TABLE 1

### MS/MS performance: results for the selective screening of PKU

<table>
<thead>
<tr>
<th>Reference</th>
<th>Chace et al., 1998 (n tested = 203)</th>
<th>Ceglarek et al., 2002 (n tested = 10,136)</th>
<th>Zytkoicz et al., 2001 (n tested = 257,000)</th>
<th>Wilcken et al., 2003* (n tested = 362,000)</th>
<th>Schulze et al., 2003† (n tested = 250,000)</th>
<th>Frazier et al., 2006† (n tested = 239,415)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>TEST</strong></td>
<td>FLUOROMETRY</td>
<td>MS/MS</td>
<td>FLUOROMETRY</td>
<td>MS/MS</td>
<td>MS/MS</td>
<td>MS/MS</td>
</tr>
<tr>
<td>Metabolic markers and cut-off values (µmol/L)</td>
<td>Phe &gt; 258</td>
<td>Phe &gt; 180 and Phe/Tyr &gt; 2.5</td>
<td>Phe &gt; 120</td>
<td>Phe &gt; 120 and Phe/Tyr &gt; 2</td>
<td>Phe &gt; 139 and Phe/Tyr &gt; 1.5</td>
<td>Phe &gt; 150</td>
</tr>
<tr>
<td>Metabolic markers and cut-off values (µmol/L)</td>
<td>Phe &gt; 157 (borderline CV) or Phe &gt; 250 (diagnostic CV) and Phe/Tyr &gt; 3.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Calculation scenarios</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Prevalence (%)</td>
<td>9.4</td>
<td>0.03</td>
<td>0.007</td>
<td>0.01</td>
<td>0.02</td>
<td>0.03</td>
</tr>
<tr>
<td>True positives (n)</td>
<td>19</td>
<td>19</td>
<td>19</td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>False positives (n)</td>
<td>91</td>
<td>3</td>
<td>1</td>
<td>141</td>
<td>125</td>
<td>27</td>
</tr>
<tr>
<td>False negatives (n)</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>True negatives (n)</td>
<td>93</td>
<td>181</td>
<td>183</td>
<td>9992</td>
<td>10,008</td>
<td>10,106</td>
</tr>
<tr>
<td>Sensitivity (%)</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>Positive predictive value (%)</td>
<td>17.27</td>
<td>86.36</td>
<td>95.00</td>
<td>2.08</td>
<td>2.34</td>
<td>10.00</td>
</tr>
<tr>
<td>Negative predictive value (%)</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>Recall rate (%)</td>
<td>54.19</td>
<td>10.84</td>
<td>9.85</td>
<td>1.42</td>
<td>1.26</td>
<td>0.30</td>
</tr>
</tbody>
</table>

n: number of subjects; MS/MS: tandem mass spectrometry; PKU: phenylketonuria; Phe: phenylalanine; Tyr: tyrosine; CV: cut-off value.
* Three subjects for whom the diagnosis could not be confirmed were included among the false positives in scenario (a) and among the true positives in scenario (b).
† Scenario (a) considers as positives subjects with two tests results above the borderline cut-off value or one test result above the diagnostic cut-off value; scenario (b) considers as positives subjects with one test result above the diagnostic cut-off value; and scenario (c) considers as positives subjects with one test result above the borderline or diagnostic cut-off value.
‡ Assumed number.
The comparison of MS/MS results based on phenylalanine quantification with those that also take into consideration the phenylalanine/tyrosine ratio shows that this ratio substantially reduces the false-positive rate and consequently improves the test’s specificity, but also, and most importantly, the positive predictive value and the recall rate (Table 1). Several authors have underscored the usefulness of this ratio in reducing the false-positive rate in neonatal PKU screening [Frazier et al., 2006; Chace and Kalas, 2005; Ceglarek et al., 2002; Zytkovicz et al., 2001; Chace et al., 1998]. Furthermore, in an analysis limited to nine neonates affected with hyperphenylalaninemia and 13 others with PKU, Ceglarek and colleagues [2002] found that the phenylalanine/tyrosine ratio permits complete discrimination between these two clinical entities. However, Frazier and colleagues [2006] reported that almost 20% of samples (number of subjects not indicated) from children receiving parenteral feeding had an elevated phenylalanine level and a high phenylalanine/tyrosine ratio, although these had normalized by the second sample. On the other hand, Zytkovicz and colleagues [2001] reported that in newborns with transiently elevated phenylalanine levels (139 to 254 μmol/L), the phenylalanine/tyrosine ratio was still under the cut-off value of 1.5. In addition, these authors found that the phenylalanine/tyrosine ratio was greater than 5 and than 1.5 in all the infants with PKU and non-PKU hyperphenylalaninemia, respectively, both on the initial and repeat tests. Even though the phenylalanine/tyrosine ratio improves MS/MS performance, the variability in results with respect to the positive predictive value and the recall rate should be noted. In fact, the results vary from 10% to 100% for the positive predictive value and from 0.005% to 9.85% for the recall rate. If the two studies representing particular cases [Frazier et al., 2006; Chace et al., 1998] are excluded, the positive predictive value would vary from 10% to 33.53% and the recall rate from 0.025% to 0.3%. Indeed, the age of the newborns in Chace and colleagues’ study [1998] was much lower than the age of sampling in most neonatal screening programs, and the phenylalanine cut-off value was much higher than in most of the other studies. In Frazier and colleagues’ study [2006], the different scenarios represent different classification approaches of the results obtained with two cut-off values, one borderline and the other diagnostic.

b) Performance results when comparing MS/MS to fluorometry

Chace and colleagues [1998] reported a high correlation between fluorometry and MS/MS results for neonatal PKU screening, with a Pearson coefficient of 0.817. If MS/MS performance is compared with that of fluorometry on the basis of the phenylalanine marker (Table 1), the results of the two studies concur, showing that both tests correctly classify affected newborns and that the use of MS/MS does not offer any gain in terms of sensitivity, although it does reduce the number of false positives and the recall rate. These advantages are however more modest in Ceglarek and colleagues’ study [2002] than in that of Chace and colleagues [1998]. This situation may be due to the use of a high phenylalanine cut-off value and to the recruitment of newborns under 24 hours of age in Chace and colleagues’ study [1998], whereas the phenylalanine concentration normally increases with age.

5.2.2.3 MS/MS performance for the selective screening of TT1

MS/MS performance data for the selective neonatal screening of TT1 were extracted from four studies, one of which was specific to this disease [Sander et al., 2006], while the other three investigated groups of inborn errors of metabolism [Schulze et al., 2003a; Wilcken et al., 2003; Zytkovicz et al., 2001]. Table 2 presents the results of these studies, indicating for each the number of true and false positives, of true and false negatives, as well as the performance criteria extracted from the article or calculated for each metabolic marker used. The number of samples analyzed by MS/MS varied from 61,344 to 362,000. The metabolic markers used to classify MS/MS results were tyrosine [Wilcken et al., 2003], tyrosine and the tyrosine/phenylalanine ratio [Zytkovicz et al., 2001], succinylacetone [Sander et al., 2006], or tyrosine and
succinylacetone together [Schulze et al., 2003a]. Succinylacetone was quantified by MS/MS in Sander and colleagues’ study [2006], and by means of a spectrophotometric test for the gamma-aminolevulinate dehydratase in the study by Schulze and colleagues [2003a], who did not indicate the cut-off value used for this assay.

It should be noted that the number of true positives was small, if not null, indicating a low disease prevalence in the studied populations. Furthermore, with one exception [Sander et al., 2006], there was a fairly high number of false positives in these studies, which translates into a low positive predictive value without significantly affecting the specificity (99.97-99.98%). Lastly, Sander and colleagues [2006], the only authors who quantified succinylacetone by MS/MS, reported a perfect specificity and positive predictive value as well as the lowest recall rate.

Sander and colleagues [2006] also obtained abnormal levels of succinylacetone on two samples from patients with a confirmed TT1 diagnosis, which were retrospectively analysed by MS/MS. In contrast, the succinylacetone level was lower than 1 µmol/L for both parents of one of these patients, which indicates that MS/MS discriminates between homozygotes and heterozygotes. Lastly, Sander and colleagues [2006] also reported (without showing supporting data) that succinylacetone levels do not correlate with tyrosine concentrations or with gestational age or birthweight. They even pointed out that none of the patients in their study would have been detected by screening if only tyrosine quantification had been used. Feuchtbaum and colleagues [2006b] did not detect any case of TT1 in the 353,894 newborns whose tyrosine levels were measured by MS/MS. Even if this observation could be explained by a low disease prevalence, the possibility of false negatives cannot be completely ruled out. In North Carolina, for example, because one patient with TT1 had not been detected by screening, the disease was excluded from the neonatal screening program [Frazier et al., 2006]. In addition, other authors have called attention to the lack of specificity of tyrosine quantification by MS/MS for neonatal TT1 screening [Comeau et al., 2004].

### TABLE 2

**MS/MS performance: results for the selective screening of TT1**

<table>
<thead>
<tr>
<th>Reference</th>
<th>Zytкович et al., 2001 (n = 164,000)</th>
<th>Wilcken et al., 2003 (n = 362,000)</th>
<th>Schulze et al., 2003a* (n = 250,000)</th>
<th>Sander et al., 2006 (n = 61,344)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Metabolic markers and cut-off values (µmol/L)</td>
<td>Tyr &gt; 442</td>
<td>Tyr &gt; 442 and Tyr/Phe &gt; 6</td>
<td>Tyr &gt; 500</td>
<td>SAC &gt; 10</td>
</tr>
<tr>
<td>Calculation scenarios</td>
<td>Scenario a</td>
<td>Scenario b</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Prevalence (%)</td>
<td>0</td>
<td>0</td>
<td>0.0006</td>
<td>0.0004</td>
</tr>
<tr>
<td>True positives (n)</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>False positives (n)</td>
<td>42</td>
<td>38</td>
<td>69</td>
<td>52</td>
</tr>
<tr>
<td>False negatives (n)</td>
<td>?</td>
<td>?</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>True negatives (n)</td>
<td>163,958</td>
<td>163,962</td>
<td>361,929</td>
<td>249,947</td>
</tr>
<tr>
<td>Sensitivity (%)</td>
<td>?</td>
<td>?</td>
<td>0</td>
<td>100</td>
</tr>
<tr>
<td>Specificity (%)</td>
<td>99.97</td>
<td>99.98</td>
<td>99.98</td>
<td>99.98</td>
</tr>
<tr>
<td>Positive predictive value (%)</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1.89</td>
</tr>
<tr>
<td>Negative predictive value (%)</td>
<td>?</td>
<td>?</td>
<td>99.99</td>
<td>100</td>
</tr>
<tr>
<td>Recall rate (%)</td>
<td>0.03</td>
<td>0.02</td>
<td>0.02</td>
<td>0.02</td>
</tr>
</tbody>
</table>

n: number of subjects; MS/MS: tandem mass spectrometry; Phe: phenylalanine; SAC: succinylacetone; TT1: tyrosinemia type 1; Tyr: tyrosine.
One subject lost to follow-up after a positive MS/MS result was considered as a false positive in scenario (a) and as a true positive in scenario (b).
5.2.2.4 MS/MS performance for the selective screening of MCADD

MS/MS performance data for the selective neonatal screening of MCADD were extracted from nine studies, four of which were specific to this disease [Andresen et al., 2001; Carpenter et al., 2001; Pourfarzam et al., 2001; Chace et al., 1997], while the other five investigated groups of inborn errors of metabolism [Feuchtbaum et al., 2006b; Frazier et al., 2006; Schulze et al., 2003a; Wileken et al., 2003; Zytkovicz et al., 2001]. Table 3 presents the results of these studies, indicating for each the number of true and false positives and of true and false negatives, as well as the performance criteria extracted from the article or calculated for each metabolic marker used. The octanoylcarnitine (C8) marker was used in all of these studies, although with different cut-off values. In some studies, this marker was used in combination with other acylcarnitines or their ratios.

The number of samples tested varied from 100,600 to 930,078. The results of the studies indicate that MS/MS has an excellent specificity (99.98-100%), sensitivity (100%) and negative predictive value (100%) for neonatal MCADD screening. However, three studies did not report on the latter two criteria, neither did they provide data to allow their calculation. On the other hand, in Carpenter and colleagues’ retrospective study [2001], 12 of the 13 archived dried blood samples from patients with MCADD proved positive on MS/MS (sensitivity of 92.31%). The authors stated that no test involving acylcarnitines or a ratio of acylcarnitines could have identified the 13th patient (homozygous for A985G) who developed an encephalopathy on the second day of life, which led to a carnitine depletion.

As with the other diseases, there was a wide variation in study results for the positive predictive value (19.23-100%), despite minor differences in disease prevalence between studies (0.004-0.008%). This variation, as well as that observed for the recall rate (0.004-0.028%), might be due to the different metabolic markers and cut-off values used. The effect of the latter on the proportion of false-positives and, consequently, on the specificity has been discussed above. A comparison of the studies by Carpenter and colleagues [2001] and Wilcken and colleagues [2003] shows the indirect effect of using different cut-off values on the positive predictive value. In fact, although the entire population of the first study was included in the second, the number of false positives was different, since the methods for classifying the screening results were not the same (Table 3). This led to a substantial difference in the positive predictive value. Although the use of metabolite ratios, especially C8/C6 and/or C8/C10, seems to reduce the number of false positives and thus improve the positive predictive value and the recall rate, this does not apply to Schulze and colleagues’ study [2003a]. These authors did not clearly indicate, however, which metabolite combination or metabolite ratio was used or if the same markers were used for all patients.

23. We decided to present the results of both of these two studies because the methods for classifying the screening results were not the same.
### Table 3

**MS/MS performance: results for the selective screening of MCADD**

<table>
<thead>
<tr>
<th>Reference</th>
<th>Chace et al., 1997 (n = 283,803)</th>
<th>Carpenter et al., 2001* (n = 275,653)</th>
<th>Andresen et al., 2001 (n = 930,078)</th>
<th>Pourfarzam et al., 2001 (n = 100,600)</th>
<th>Zytkovicz et al., 2001a (n = 184,000)</th>
<th>Schulze et al., 2003a (n = 250,000)</th>
<th>Wilcken et al., 2003 (n = 362,000)</th>
<th>Feuchbaum et al., 2006b (n = 353,894)</th>
<th>Frazier et al., 2006 (n = 239,415)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Metabolic markers and cut-off values (CV) (µmol/L)</strong></td>
<td>C8 &gt; 0.3 C8/C10 &gt; 2 C8/C2 &gt; 0.1</td>
<td>C8 &gt; 0.8 (first test) C8 &gt; 1.0 (second test)</td>
<td>C6, C8, C10, C10:1 (CV not indicated)</td>
<td>C8 &gt; 0.3 and C8/C6 &gt; 4.0</td>
<td>C8 &gt; 0.5 {C6 &gt; 0.21 and/or C8 &gt; 0.32}, C10:1 &gt; 0.28, C10 &gt; 0.48, C8/C2 &gt; 0.02, C8/C10 &gt; 1.6, C8/C12 &gt; 1.6</td>
<td>C8 &gt; 1.0</td>
<td>Not indicated</td>
<td>C8/C10 &gt; 3 (≥ C6 &gt; 0.63, C10:1 &gt; 0.31)</td>
<td></td>
</tr>
<tr>
<td>Prevalence (%)</td>
<td>0.006</td>
<td>0.004</td>
<td>0.006-0.007</td>
<td>0.008</td>
<td>0.005</td>
<td>0.006</td>
<td>0.005</td>
<td>0.004</td>
<td>0.007</td>
</tr>
<tr>
<td>True positives (n)</td>
<td>16</td>
<td>11 or 12†</td>
<td>53-62‡</td>
<td>8</td>
<td>8</td>
<td>10</td>
<td>16</td>
<td>17</td>
<td>13</td>
</tr>
<tr>
<td>False positives (n)</td>
<td>0</td>
<td>12 or 11†</td>
<td>0-9‡</td>
<td>6</td>
<td>0</td>
<td>42</td>
<td>46</td>
<td>6</td>
<td>2</td>
</tr>
<tr>
<td>False negatives (n)</td>
<td>0†</td>
<td>?</td>
<td>?</td>
<td>0</td>
<td>0</td>
<td>?</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>True negatives (n)</td>
<td>283,787</td>
<td>275,630</td>
<td>930,016</td>
<td>100,586</td>
<td>100,592</td>
<td>183,948</td>
<td>249,938</td>
<td>361,977</td>
<td>353,879</td>
</tr>
<tr>
<td>Sensitivity (%)</td>
<td>100</td>
<td>?</td>
<td>?</td>
<td>100</td>
<td>100</td>
<td>?</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>Positive predictive value (%)</td>
<td>100</td>
<td>47.83-52.17†</td>
<td>85.48-100‡</td>
<td>57.14</td>
<td>100</td>
<td>19.23</td>
<td>25.81</td>
<td>73.91</td>
<td>86.67</td>
</tr>
<tr>
<td>Negative predictive value (%)</td>
<td>100</td>
<td>?</td>
<td>?</td>
<td>100</td>
<td>100</td>
<td>?</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>Recall rate (%)</td>
<td>0.006</td>
<td>0.008</td>
<td>0.007</td>
<td>0.014</td>
<td>0.008</td>
<td>0.028</td>
<td>0.025</td>
<td>0.006</td>
<td>0.004</td>
</tr>
</tbody>
</table>

* The performance criteria calculation is based on a cut-off value of C8 ≥ 0.8. Note that the population of this study was included in that of Wilcken and colleagues [2003]. We none-the-less decided to present the results of both of these studies, since the methods for classifying the screening results were not the same.

† The true or false positive status of one patient was unclear, because he had a positive MS/MS result and very low medium- and long-chain fatty acid oxidation levels, but an intermediary level of enzyme activity consistent with heterozygous status. The authors report a single copy of the A985G mutation, but it is not clear whether the second allele was normal or whether the patient was a composite heterozygote (i.e. if the second allele carried a mutation other than A985G). The authors considered this patient a carrier at low risk for developing symptoms.

‡ The true or false positive status of nine patients was unclear. These include: 1) three subjects for whom the false positive status was highly probable, given that their respective genotypes were A985G/A351C, A985G/T489G and 0/C734T. Indeed, the authors considered the A351C, T489G and C734T mutations as silent because the metabolic profiles of these patients were moderately disturbed and normalized with time, except for the patient with the T489G mutation. In addition, the metabolic profile of the patient carrying the A351C mutation was normal, even during an episode of high fever and reduced oral feeding; 2) three subjects in whom sequencing was not performed for reasons not stated in the article; 3) two subjects in whom a single copy of the A985G mutation was identified and whose metabolic profiles remained profoundly disturbed, possibly because of another inborn error of metabolism; and 4) one subject in whom no mutations were detected and whose metabolic profile remained profoundly disturbed, possibly because of another inborn error of metabolism.

§ Assumed number.
The differences in results for the positive predictive values and recall rates may also be related to other analytical parameters or to specific characteristics of the study populations. In fact, Pourfarzam and colleagues [2001] reported that, of the six false positives identified on the basis of an elevated C8 level, five had been born prematurely. Similarly, of the 52 newborns with a positive MS/MS result in Zytkovicz and colleagues’ study [2001], 25 (48%) were hospitalized in the intensive care unit or had a low birthweight. Based on data from the same neonatal screening program, Comeau and colleagues [2004] estimated the positive predictive value for newborns with a birthweight greater than 2500 g to be 26%, while Zytkovicz and colleagues [2001] estimated it to be 19%24. Carpenter and colleagues [2001], on the other hand, reported that the C8 levels did not vary substantially with the birthweight or the age at sampling in their study population (data not shown).

It should also be mentioned that several scenarios were considered for the calculation of the performance criteria for two studies specific to MCADD [Andresen et al., 2001; Carpenter et al., 2001]. For Carpenter and colleagues [2001], the true- or false-positive status of one patient was not totally clear, because he had a positive MS/MS result and very low levels of medium- and long-chain fatty acid oxidation, but an intermediate level of enzyme activity, which is consistent with a heterozygous status. The authors reported that the patient had a single copy of the A985G mutation, but it is not clear if the second allele was normal or if it carried a mutation other than A985G. They considered this patient a carrier at low risk for developing symptoms. For Andresen and colleagues [2001], the genotype results did not clearly show the true- or false-positive status of nine patients. These included: 1) three subjects for whom sequencing was not performed for reasons not stated in the article; 2) two subjects in whom a single copy of the A985G mutation was identified and whose metabolic profiles remained markedly disturbed, possibly because of another inborn error of metabolism; 3) one subject in whom no mutations were detected and whose metabolic profile remained markedly disturbed, possibly because of another inborn error of metabolism; and 4) three subjects who may have been false positives, based on their genotype (A985G/A351C, A985G/T489G and 0/C734T, respectively). The authors considered the A351C, T489G and C734T mutations as silent because the associated metabolic profiles were moderately disturbed and normalized with time (with the exception of the patient with the T489G mutation). In addition, the metabolic profile of the patient carrying the A351C mutation was normal, even during an episode of high fever and reduced oral feeding. In order to take into account all possible situations, the two extreme scenarios are considered in Table 3 for each of these studies.

5.2.3 Summary

Our review of the evidence on MS/MS performance confirms the reservations expressed in the previous systematic reviews with respect to the quality of the available studies. These reservations specifically concern the study designs, the reporting of results, the population selection procedures and the lack of standardization with regard to the diagnostic confirmatory tests and to several factors that could affect the quality of MS/MS analyses [Marsden et al., 2006; Chace et al., 2005; CDC, 2001]. Most of the available evidence derives from prospective cohort studies carried out in the context of neonatal screening programs because of the difficulty in conducting controlled prospective studies with a suitable control group, as pointed out by several authors [Shortland, 2004; Wilcken et al., 2003]. In the

---

24. The study populations of Comeau and colleagues [2004] and Zytkovicz and colleagues [2001] overlapped. The population in the former included more subjects, since it was conducted over a longer period of time (1999-2003) than the latter (1999-2001).
context of these newborn screening programs, the reference tests for confirming the diagnosis are performed only for patients with a positive MS/MS result. Consequently, the data on the proportion of false negatives are of dubious quality.

Overall, the results of the studies that were reviewed indicate that the sensitivity, negative predictive value and specificity of MS/MS are high, whether for the neonatal screening of groups of diseases or for the selective screening of PKU, MCADD and TT1. However, most of the studies did not report false negatives and did not involve a long observation period or use rigorous methods to ensure that they did not miss any cases of inborn errors of metabolism. The estimates of the sensitivity and negative predictive value of MS/MS may thus have been overestimated as a result of the limited quality of the study designs used. The positive predictive values and recall rates vary substantially from study to study. These differences might be due to the methodological differences mentioned earlier and, especially in the case of the positive predictive value, to differences in the prevalence of inborn errors of metabolism from one population to another.

Some aspects of the analytical protocols are to be noted for each of the three diseases of interest in this report. First, as pointed out by a number of authors, the benefits of MS/MS-based neonatal PKU screening derive from the simultaneous analysis of phenylalanine and tyrosine allowing for the calculation of the phenylalanine/tyrosine ratio, [Frazier et al., 2006; Chace et al., 2005; 1998; Ceglarek et al., 2002; Zytkovicz et al., 2001]. For TT1, the use of MS/MS is conceivable only if succinylacetone is quantified in addition to tyrosine. This is because tyrosine quantification by MS/MS, as by other screening techniques, is neither sensitive [Feuchtbaum et al., 2006b; Roscher and Olgemoller, 2004; Wilcken et al., 2003] nor specific enough [Comeau et al, 2004; Wilcken et al., 2003; Zytkovicz et al., 2001]. In Quebec, these limitations have been recognized for a long time, and neonatal TT1 screening has been based on succinylacetone quantification, first as a second tier test (1980-1997) for neonates with elevated tyrosine levels and, since 1997, as a first tier test [Laflamme et al., 2006; CETS, 1998]. However, succinylacetone quantification by MS/MS has been evaluated by only one study [Sander et al., 2006] and raises organizational issues, since it requires an additional extraction step and a separate analysis from that for the other metabolites. These issues could be resolved by adapting a recently proposed method, which still needs to be validated [Allard, 2005]. Lastly, for MCADD, the use of the C8/C2, C8/C6 and/or C8/C10 ratios in addition to the metabolite C8 is essential since it improves the specificity and positive predictive value of MS/MS [Frazier et al., 2006; Chace et al., 2005; 1997; Pourfarzam et al., 2001]. However, Zytkovicz and colleagues [2001] indicated (data not shown) that the C8/C10 ratio might be more useful for patients who are homozygous for the A985G mutation than for composite heterozygotes. These ratios may nonetheless be useful for the differential diagnosis between MCADD and other inborn errors of metabolism characterized by an elevated level of the metabolite C8, such as multiple acyl-CoA dehydrogenase (MAD) deficiency, also called glutaric aciduria type II (GAIi), and medium/short-chain L-3-hydroxyacyl-CoA dehydrogenase deficiency (M/SCHADD). Therefore, the optimal choice of metabolic markers for neonatal MCADD screening is still open for debate [Marsden et al., 2006; Chace et al., 2005].

Lastly, the performance results should be interpreted as indicators of MS/MS validity, that is, the ability to correctly classify newborns on the basis of the investigated biochemical abnormalities. The number of true positives identified in neonatal screening programs is not directly equivalent to the number of infants who will need treatment for life. All infants identified as positives should, however, undergo a diagnostic

25. Very few authors have attempted to determine, for all inborn errors of metabolism, the proportion of neonates identified by MS/MS who are likely to benefit from this screening in light of the timing of screening, the need for an early treatment and the impact of such treatment on the prognosis [Schulze et al., 2003a].
workup and be followed by a specialized team. Although preventive measures are proposed for all infants with TT1 or MCADD, a distinction needs to be made for those with hyperphenylalaninemia. Infants with PKU must be put on a restricted diet without delay and closely followed up, whereas girls with non-PKU hyperphenylalaninemia will not benefit from the information on their status until they reach child-bearing age, for the monitoring of their own pregnancies. For MCADD, the clinical benefit documented thus far mainly concerns infants with the most severe forms characterized by acute metabolic decompensation (provided they are identified before the onset of symptoms). The proportion that these infants represent among all identified cases is estimated at approximately 75%, but this figure needs to be confirmed [Grosse et al., 2006b]. The major difference, in comparing this situation with that of infants with hyperphenylalaninemia, is that the current knowledge of the natural history of MCADD and of its genotype/phenotype correlation is insufficient to establish a prognosis at a young age and to target the clinical management accordingly. The situation would be even more complex if screening were to be expanded to all inborn errors of metabolism, considering that the natural history of many of these diseases is poorly known.
The objective of this section is to examine, in light of the available literature, the costs, cost-effectiveness and cost-utility of MS/MS-based neonatal blood screening particularly for PKU, MCADD and TT1. It is also aimed at providing budgetary information on certain capital and operating costs relevant to this type of screening.

6.1 Literature review

The table in Appendix K presents a synthesis of the literature reviewed for this analysis. Only articles meeting the following criteria are presented: screening context potentially applicable to Quebec, inborn errors of metabolism clearly defined, robust methodology and results, and articles published after 2001. According to these criteria, no economic evaluation of neonatal screening in the Quebec context was carried out. However, one evaluation in the Canadian context that is useful for the purposes of our analysis was identified, namely that of Tran and colleagues [2006]. These authors evaluated the cost-effectiveness of MS/MS-based MCADD screening as compared to that of clinical diagnosis (in the absence of screening) using literature data and the recent experience (2005) of the Nova Scotia screening program. The unit cost of screening was determined to be C$2.40. Based on a decision-tree model and a sensitivity analysis, the authors concluded that the benefits of an MS/MS-based neonatal screening are higher than those of clinical diagnosis. The cost-effectiveness of screening is mainly related to the reduction in the cost of medical care during the infants’ lifetime. The authors also presented budgetary impact estimates for a 5-year period for the cohort of Nova Scotia newborns (without sensitivity analysis). They stressed the limitations of their analysis, in particular, the fact that the purchase cost of the equipment was entirely allocated to the first fiscal year. It is worth noting that a single MS/MS instrument was considered at a very low purchase cost (C$200,000).

The cost-effectiveness of MS/MS-based neonatal MCADD screening was also demonstrated by Venditti and colleagues [2003]. However, their result is based on the assumption that the operating costs of MS/MS are already covered by its use for PKU screening. A systematic review by the Medical Advisory Secretary [MAS, 2002] concluded that an expanded MS/MS-based neonatal screening program would be cost-effective and stated that the inclusion of other inborn errors of metabolism in such a program would not generate any additional cost. The results of three other studies [Autti-Rämö et al., 2005; Insinga et al., 2002; Schoen et al., 2002] also concluded that MS/MS-based screening is cost-effective for a number of inborn errors of metabolism. Lastly, Pandor and colleagues [Pandor et al., 2006b; 2004] showed, by economic modelling, that substituting MS/MS for the existing technology for PKU screening alone is not justified, but that MS/MS-based screening would be cost-effective if at least one other inborn error of metabolism, such as MCADD, were added.

There was a wide variability between the studies identified in the economic literature with regard to the inborn errors of metabolism considered, the probabilities of neurological disabilities and death in unscreened children with MCADD, the incidence data, and the chosen measures of efficacy. Outcomes considered for evaluating efficacy could include medical complications, hospitalizations, care and treatment, moderate or severe neurological disabilities, and prevented deaths.
the specific case of MCADD screening, the cost-effectiveness mainly relates to the avoidance of treatment and hospitalization costs (paediatric intensive care unit).

6.2 Cost estimation method

A budget impact approach was used to estimate the cost of MS/MS-based neonatal screening of PKU, MCADD and TT1. It must be recalled that MCADD cannot be detected by means other than MS/MS. The cost estimates do not derive from the direct observation of the use of this technology, but rather from various validated sources of information (see Section 6.2.1). The analysis is therefore aimed only at providing budgetary information on certain capital and operating costs for a laboratory based on the scarce data that are available.

Incremental costs were estimated in as far as possible, despite the difficulties encountered in reconstituting the costs of current screening practice. Equivalent annual incremental costs (EAICs) were then estimated. These allow the considerable investment costs related to the purchase of the MS/MS and ancillary equipment, the laboratory facility installation, and the training to be spread out over several years, while taking into account the opportunity cost that these investments involve. The EAICs thus represent the annual value of resources used for MS/MS-based neonatal screening. The cost estimates were determined from market prices and are expressed in 2006 Canadian dollars.

The estimates presented are limited to the costs generated by the main components of the MS/MS-based screening: the costs relating to the laboratory facility installation; the purchase cost of the two MS/MS instruments required, including a basic service contract, the ancillary equipment, the operator training, and the supplies; and the costs covering the time of professionals involved in such screening (two laboratory technicians). The costs covering the time of nurses (sample collection) and of a computer technician (data entry) are not taken into consideration, since they are substantially the same as those spent for the current screening method. The cost associated with the clinical biochemist’s work (for the analysis of results and formulation of recommendations) could not be estimated because of the difficulty in evaluating what proportion of his/her time would be devoted to MS/MS-based screening. Indeed, this time depends on the roles and responsibilities of each professional involved in screening. It will be noted, however, that the hourly rate for an experienced clinical biochemist is at least CA$53.25 (including benefits) and, to our knowledge, the neonatal screening laboratory presently employs one full-time clinical biochemist.

The cost of the laboratory facility installation for MS/MS-based neonatal screening will have to be incurred every 10 years or so. The same holds true for the ancillary equipment and the bar code reader. According to specialists, both MS/MS instruments should be reserved exclusively for neonatal screening and operate concurrently, so as to avoid any interruption of screening in the event of technical problems. Even if all service contracts are established for a maximum of five years, the average lifespan of an MS/MS instrument can, according to some specialists, be as long as 7 to 10 years.

---

27. The EAIC is the current value of the cost generated by MS/MS-based screening (3% discount rate) expressed on an annual basis.
28. This concerns the installation of hoods, laboratory tables, etc.
29. The two tandem mass spectrometers include the injector, carrousel, quaternary and/or binary pumps, valves, filter, column heater, MS/MS detector, operating software and research database (one for the two instruments), computer, screen and printer.
30. The ancillary equipment includes a system for both mass spectrometers, namely, a compressor, a nitrogen generator and an uninterruptible power supply (UPS).
31. The supplies include reagents, multiwell plates and multichannel pipettor tips.
since the instrument consists of modules that can be changed separately, if need be. We therefore considered that the average cost of purchasing two mass spectrometers and of training would have to be incurred every eight years on average (a minimum of 5 years and a maximum of 10 years), whereas the average cost of supplies, service contracts for instrument maintenance, and personnel time would have to be incurred every year. Most specialists in the field consider the latter costs relatively low.

The costs generated by repeat tests on the same sample or on a second sample (performed in case the repeat test is positive) and by the diagnostic confirmatory tests were omitted. Similarly, the tests for succinylacetone were not taken into account. Lastly, the benefits of early screening, such as the prevention of severe handicaps and the avoided costs of medical care, were not estimated.

6.2.1 Sources of information

Nearly all of the cost estimates were derived from information provided orally and in writing by specialists in the field (a biochemist in a department of medical genetics, laboratory technicians, and a chief technologist). Face-to-face meetings as well as communications via e-mail or telephone were undertaken in order to clarify certain issues. Information from the industry also helped confirm certain cost data. Lastly, the hourly rates for the different categories of personnel that are involved in screening were extracted from the Conseil du trésor’s website.

6.2.2 Results

Table 4 shows the equivalent annual incremental cost (EAIC) of each of the main components of an MS/MS-based neonatal screening program for PKU, TT1 and MCADD. Other inborn errors of metabolism could be added later without substantially increasing the cost (it must be recalled that the instrument can actually screen for more than 30 diseases). The table also displays at which frequency the cost of these components would have to be incurred. All of these data were validated with specialists in the field.

The results in Table 4 show that the annual cost of an MS/MS-based neonatal screening program run in one laboratory in Quebec would be about CA$255,231. The main expenditures would be for the acquisition of the MS/MS and ancillary equipment. The cost of laboratory technicians’ time is also substantial, since they should ideally work on an annual basis of 365 days so that screening is not interrupted. Two scenarios were considered on the basis of the opinion of the specialists consulted: 1) 1.5 full-time equivalents (FTEs); and 2) 2 FTEs. The costs would thus vary from $91,460 to $121,946, with an estimated average cost of $106,703. The cost of the reagents (see Table L-1 in Appendix L) was estimated on the basis of an average of 75,000 tests a year, i.e., the approximate annual number of births in Quebec, although a portion of these reagents may also be used in current screening practices.

More specifically, the annual implementation and equipment cost generated by MS/MS-based screening is estimated to be about $100,569 ($149,902 if the MS/MS instruments are replaced every five years and $84,172 if they are replaced every ten years).


33. As mentioned in note 18, information obtained at the time this report was in press [MFACF, 2007] suggests that this figure had increased to 82,500 in 2006.
every 10 years). The annual operating and upgrading costs are estimated to be $47,959, with expenses related to the service contract and reagents accounting for most of this amount. It is noteworthy that the cost of an extended service contract (including a kit with on-site replacement parts) can be as high as CA$225,000 for five years, which would substantially increase the total cost of screening.

6.3 Summary

For the tandem mass spectrometry screening programs implemented to date, the literature is scarce with respect to formal evaluations of their economic aspects. According to Tran and colleagues’ results [2006], the estimated unit cost of MS/MS-based neonatal MCADD screening is low (CA$2.40), but these results raise a number of questions. In addition, the authors considered purchasing a single instrument at the surprisingly low price of $200,000. However, the results of their sensitivity analysis support the cost-effectiveness of such screening, with a unit cost of up to CA$5.00. In Quebec, where two instruments would be necessary and could cost approximately CA$325,000, the estimated unit cost would be about CA$3.4034, which supports the cost-effectiveness of MS/MS-based screening for MCADD. It will be recalled that this unit cost does not include expenses generated by repeat tests on the same sample or on a second sample (performed if the repeat test on the same sample is positive) or by the diagnostic confirmatory tests neither does it cover the work of the clinical biochemist, which varies according to the number of hours devoted to MS/MS-based screening. If, for instance, the clinical biochemist were to devote the equivalent of two and a half days a week (7 hours and 45 minutes/day) to MS/MS-based screening, the annual cost of this professional would be about CA$35,704 (hourly rate of $53.25 for 36 weeks). If he/she worked full-time on MS/MS-based screening, this figure would be CA$71,408, which would increase the unit cost from CA$3.40 to CA$3.88 and CA$4.36, respectively.

---

34. This estimate is based on the assumption of 75,000 births a year in Quebec. The estimated total EAIC ($255,231) divided by this number of births (75,000) yields a unit cost of CA$3.40. If the latest statistic for the number of births in 2006 (82,500) is used, the cost decreases to CA$3.09.
<table>
<thead>
<tr>
<th>COST CATEGORIES FOR MS/MS-BASED SCREENING</th>
<th>ESTIMATED COST</th>
<th>EAIC *</th>
<th>TOTAL EAIC</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Installation and equipment costs</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fixed costs</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Laboratory facility installation (hoods, laboratory benches, etc.)</td>
<td>$20,000†</td>
<td>$2,345 every 10 years</td>
<td></td>
</tr>
<tr>
<td>Equipment (2 tandem mass spectrometers‡)</td>
<td>$650,000§</td>
<td>$92,597 (if every 8 years); ($141,930 if every 5 years; $76,200 if every 10 years).</td>
<td></td>
</tr>
<tr>
<td>Ancillary equipment (compressor, nitrogen generator and UPS)</td>
<td>$45,000</td>
<td>$5,275 every 10 years</td>
<td></td>
</tr>
<tr>
<td>Bar code reader</td>
<td>$3,000</td>
<td>$352 every 10 years</td>
<td></td>
</tr>
</tbody>
</table>

| Operating and upgrading costs           |                |        |            |
| Fixed costs                             |                |        |            |
| Cost of basic on-site training† (of 4 days’ duration for about 5 operators) | $12,000 | $1,709 (if every 8 years) ($2,620 if every 5 years; $1,407 if every 10 years) | |
| Basic service contract for equipment maintenance, including a 1-year warranty (between $100,000 and $225,000 for 5 years) | $20,000 to $45,000 | $20,000 to $45,000 every year | |

| Variable costs                          |                |        |            |
| Reagents (internal standards, methanol, butanol, aluminium, acetonitrile), multiwell plates, and multichannel pipettor tips¶ | $26,250 | $26,250 every year | |

| Cost of personnel**                    |                |        |            |
| Variable cost                          |                |        |            |
| Laboratory technician**                | $106,703 ($91,460; $121,946) | $106,703 every year | |

| Estimated total EAIC                  |                |        | $255,231 |

* EAIC: equivalent annual incremental cost calculated with a 3% discount rate [Drummond et al., 2005]. These EAICs are the annual operating costs form a budget impact perspective for the hospital.
† These costs may vary according to the complexity of the building and could range from $10,000-$50,000 (2006 Canadian dollars) (according to one expert in the field).
‡ A second MS/MS is essential to avoid the interruption of screening during downtime of the first instrument.
§ After allowing for a 5% partial refund of the GST and QST, net taxes are included in the purchase cost.
¶ Training is provided when the new equipment is purchased only.
† The total cost of the reagents, pipettor tips and multiwell plates works out to about $0.35/sample x 75,000 births.
** See Table L-2 in Appendix L.
++ Two scenarios were considered: 1) 1.5 FTEs; and 2) 2 FTEs. Since the hourly rate is $31.47, we have, in scenario 1: 1.5 x $31.47/hr x 7 hrs 45 min x 250 days = $91,460; and, in scenario 2: $121,946.
Many ethical, psychosocial and organizational issues relating to the use of MS/MS for neonatal screening are raised in the literature. These issues, discussed below by topic, pertain to the impact of the uncertainty associated with neonatal screening, the impact of the knowledge generated by neonatal screening, the risks of stigmatization, the consent procedures, the gaps in professional knowledge, societal impacts, the interface between research and screening programs, and organizational aspects. Appendix M consists of a table summarizing the main issues raised and the corresponding references. A more in-depth discussion of the concerns expressed in the literature is provided in the technical report [Makni et al., 2007]. Most of the issues raised are relevant regardless of the technology used for neonatal screening. Only a few issues are accentuated by the use of MS/MS for the screening of the three diseases of interest.

If we consider the introduction of MS/MS-based neonatal screening specifically for the three diseases of interest, the following issues are likely to take on special significance:

1) the impact of uncertainty, with, on the one hand, possibly decreased numbers of false positives for PKU, and, on the other, more newborns being diagnosed with MCADD, with no clear knowledge of its penetrance or long-term prognosis;

2) the consent, for which the optimal procedure and its implications for health care service provision have not been determined;

3) the provision of services, with an increased need for genetic counselling and for follow-up by metabolic disease specialists for the additional MCADD cases identified;

4) the education of health professionals, with an increased need for primary care professionals trained to provide information on MCADD prior to screening and to follow patients identified as newborns with MCADD.

More acute problems are raised, however, with the possibility of expanding screening to a large number of diseases, in particular, when their natural history is less clear and the benefits of early treatment are less well established. In this regard, some authors favour maintaining the conventional approach put forth in the criteria proposed by Wilson and Junger [1968], which places the newborn’s interest at the centre of the evaluation of the benefits and risks of screening. Others feel that these criteria need to be expanded in order to take into consideration the familial benefits in terms of informed reproductive choices. Given the diversity of the phenotypic expression of the diseases and the variable penetrance of mutations, it is important to evaluate each

---

35. With respect to consent, there is a continuum in the degree of parental choice, the type of information provided to parents, the method used to convey this information, and the manner in which parental decision is expressed and recorded in the medical chart [Hargreaves et al., 2005]. In Quebec, consent for newborn screening is inferred from the general consent to care and services signed by the parturient upon hospital admission.
inborn error of metabolism separately [Holtzman, 2003]. However, some authors propose a more global approach, since neonatal screening for one of these diseases may reveal another disease not targeted by the screening program [Pollitt, 2006]. This possibility raises challenges with respect to the management of results. Once again, opinions differ\(^\text{36}\), with those who consider this problem as one of the main ethical issues associated with expanding neonatal screening opposing those who consider that this problem already exists with PKU screening, which occasionally detects disorders of BH4 metabolism [Pollitt, 2006; Matern, 2002]. These debates contribute to the diversity of procedures adopted in different countries. This diversity could also be explained by the differences in the organization and funding of neonatal screening services, the questionable quality of the data derived from the available literature, the lack of standardisation of neonatal screening criteria, and the diversity in professional affiliations among members of the committees that evaluate these criteria [Pollitt, 2006].

The following approaches are recommended for dealing with the issues raised by implementing MS/MS and expanding neonatal screening, whether, initially, for MCADD or for a wider range of diseases: improving communication strategies with patients, bolstering efforts to train health professionals, and organizing neonatal screening as an integrated system of services including a comprehensive monitoring system with data collection on patients’ health status and on service utilization. These approaches are linked in that an improvement in health professionals’ knowledge would, in the short term, result in better patient information and reduce accordingly the scope of the issues relating to uncertainty, knowledge and consent. In the long run, an integrated newborn screening system would lead to an increased knowledge of the diseases and a reduction in the uncertainty surrounding the interpretation of results and the prognosis. More-complete and better-quality information would then be available to patients, health professionals and health-care system administrators. The challenges associated with the development of an integrated neonatal screening system include regional legislative constraints, information protection and confidentiality issues, the funding and long-term maintenance of the equipment, the availability of the necessary expertise and financial and human resources, as well as the political will and the ability to achieve a common vision of the objectives and of the means to be put in place.

\(^{36}\) The diversity of positions between countries can be illustrated by the situation in the United States and Germany. Most of the American states have decided to require communication of results for all detected diseases. In contrast, Germany, which had undertaken screening of a large number of inborn errors of metabolism in the context of pilot projects, limited this number to 10, following an evaluation performed in 2002 by the Interdisciplinary Screening Commission of the German Society of Pediatrics and approval by the Federal Ministry for Health and Social Security. The ministry also prohibited the screening of any other inborn error of metabolism and “decreed that accidentally obtained results arising from the allowed screens must be ignored and not communicated to anyone” [Pollitt, 2006].
A part from considerations related to the technology’s performance, the decision to include a given disease in a neonatal screening program is essentially based on the ability to favourably alter the prognosis by early detection and intervention.

For PKU, the utility of neonatal screening is recognized. Even if the quality of the initial studies was not of the highest standard, the experience accumulated over time has confirmed the relevance of screening. Furthermore, screening for this condition has proven to be cost-effective. Judging from the two studies that directly compared fluorometry and MS/MS for PKU screening, MS/MS quantification of the phenylalanine/tyrosine ratio would reduce the number of false positives. Technology replacement would therefore decrease the number of newborns who would need to undergo diagnostic confirmation, and thus reduce parental anxiety as well as costs. However, as mentioned in the Institut national de santé publique du Québec (INSPQ) report, the current screening method has a sensitivity of 100% and a specificity of 99.998%, which compares favourably with the literature data. In these circumstances, the advantage of MS/MS over fluorometry is not as obvious as indicated in the literature and is not a sufficient argument for technology replacement if only PKU is considered.

For TT1, neonatal screening was introduced in the 1970s because of a founder effect in Quebec, although the efficacy of early treatment following screening had not been demonstrated. One ongoing study investigating the follow-up of more than 300 affected children tends to show a much better prognosis since the introduction of NTBC. However, it is not a prospective, controlled study since, at each period in time, these children received all available treatments because of their seemingly promising effects. The conclusions in favour of the efficacy of NTBC therapy are therefore based on comparisons between cohorts of children managed over different time periods. Furthermore, this study includes affected children who were either identified by screening or diagnosed clinically. As a result, the benefits of neonatal screening and early treatment with NTBC may be underestimated. Quebec data are, however, available, although they have not yet been published. These preliminary results suggest that the number of liver transplants has decreased substantially since the introduction of NTBC. The number of deaths associated with the disease has also decreased significantly. In light of all of these results, the clinical utility of neonatal TT1 screening is not questioned.

In view of the technology’s performance, the use of MS/MS for TT1 screening is conceivable only if succinylacetone is quantified in addition to tyrosine. However, experience with MS/MS-based TT1 screening is limited, worldwide, especially with regard to succinylacetone quantification. In fact, the clinical validity of MS/MS-based succinylacetone quantification has been evaluated in only one study [Sander et al., 2006]. Furthermore, this approach raises organizational issues, since an additional extraction procedure is required and since the succinylacetone assay cannot be performed at the same time as the analysis of amino acids and acylcarnitines. Preliminary work aiming at developing a simultaneous analysis of all of these metabolites is underway, but the performance of such an assay will need to be confirmed. Quebec would undoubtedly be a favourable environment for carrying out validation studies, ideally both analytical and clinical.

In the case of MCADD, MS/MS is the only technology that can be used for neonatal screening. Before the advent of MS/MS, only the symptomatic forms were diagnosed.
and managed. When a diagnosis was made following the onset of complications, the mortality associated with metabolic decompensation was high and the neurological sequelae were relatively frequent. For survivors and individuals with somewhat less severe forms, the avoidance of fasting seemed to have a preventive effect on the development of complications. However, the variability in the disease’s evolution makes it difficult to evaluate the efficacy of its management. With the implementation of MS/MS-based MCADD screening programs, data are starting to accumulate on the prognosis of patients who are identified and managed before the onset of symptoms. The proportion of patients who experience metabolic crises at a young age and the number of deaths were found to be lower in newborns identified by neonatal screening than in clinically diagnosed children. However, such comparisons are prone to bias, because of differences in disease spectrum between these two groups. Studies on the genotypic status of patients identified through screening tend to lend support to the existence of such differences because some of these patients seem to have a less severe form of the disease than clinically diagnosed patients. Knowledge of the genotype/phenotype correlation remains, however, very limited. Consequently, it is difficult to draw firm conclusions about the benefit of screening for all identified MCADD patients on the basis of the follow-up of clinically diagnosed patients. For the more severely affected patients, the benefit of presymptomatic management has, nonetheless, been better documented. Simple and safe preventive measures can radically alter the prognosis of these children if appropriate and timely care is provided. With the available evidence and despite its limitations, the balance between the pros and cons is in favour of neonatal screening. Indeed, the benefits of early treatment for severely affected patients are so substantial that they seem to outweigh the uncertainty surrounding the benefit for less severely affected patients. In addition, the sensitivity and specificity of MS/MS-based neonatal screening appear to be very high, especially if the C8/C6 and/or C8/C10 ratios are used in addition to the C8 metabolic marker. However, the identification of false negatives is even more difficult than for the other diseases because of a probable underdiagnosis of early lethal forms and because of variants that remain asymptomatic until adulthood. Lastly, a number of modelling exercises, which admittedly have their limitations, provide an additional argument, namely that of the cost-effectiveness of screening.

Overall, as far as the clinical utility for patients and their families is concerned, neonatal screening is justified for the three diseases of interest, despite the gaps in the evidence and the various issues that each disease raises. For MCADD, neonatal screening necessarily requires MS/MS, whose performance is one of the best for this disease. However, it will be necessary to collect data in an ongoing fashion so as to periodically re-evaluate the benefits of neonatal screening for this inborn error of metabolism. For PKU, the use of MS/MS would not substantially improve the level of neonatal screening performance currently observed in Quebec but would probably not compromise it either. If MS/MS were to be used for MCADD screening, the technology transfer for PKU would avoid duplication of analytical steps and would probably be less expensive than continuing with the current analytical method alongside MS/MS. New approaches to neonatal TT1 screening based on both tyrosine and succinylacetone quantification seem promising but must still be validated.

The relevance and optimal timing of implementing MS/MS-based screening in Quebec depend on a number of ethical, social, legal, economic and organizational issues, in addition to scientific and technical considerations. Some of these issues are discussed here, while others are beyond the scope of this report. Therefore, three separate scenarios are proposed below, together with a discussion of their respective advantages and disadvantages.
The three proposed scenarios are: 1) conducting a pilot study on screening for these three diseases over several years; 2) postponing MS/MS implementation until after validation studies on the analytical protocol combining all the analytical markers have been completed; and 3) introducing MS/MS-based screening for PKU and MCADD while, for TT1, either maintaining the current methods until the results of the above-mentioned validation studies are available or undertaking gradual technology replacement. The advantages and disadvantages of each option and the issues raised are discussed below and summarized in Appendix N.

1) Conducting a pilot study would require a clear statement of objectives, a rigorous study design, and detailed planning. Such a project would entail considerable costs, and its impact on the functioning of the current neonatal screening program would need to be discussed. It should be noted that the term «pilot project» covers a wide range of study design options. Thus, a minimalist interpretation would be a pre-implementation pilot phase with the concurrent use of two screening techniques, namely MS/MS and the conventional techniques for PKU and TT1. The advantages of this option would include a field evaluation of the costs of both MS/MS and the current screening methods; a comparison of the performance of the two analytical methods for PKU and TT1; and a collection of epidemiological, genetic and clinical data on MCADD. The epidemiological and genetic data would contribute to the advancement of knowledge about this disease, but the same would not necessarily be true for the clinical data. This is because such a study would be prone to the same limitations as research conducted elsewhere in the context of neonatal screening programs. At the other end of the spectrum, a controlled pilot study could be planned with a control group that would not undergo MS/MS-based neonatal screening but would benefit from a comparable organization of the diagnostic services. Apart from the above-mentioned advantages, a controlled pilot study would permit an evaluation of the clinical benefits of neonatal MCADD screening. Conducting such pilot studies was proposed in 1997 by the authors of the British systematic reviews, but they were not implemented until recently, mainly because of budgetary constraints. In March 2004, a prospective study37 was undertaken in the United Kingdom with the aim to screen 700,000 newborns, that is about half of all births during a 2-year recruitment period. In parallel, a systematic surveillance would be conducted during four to five years in order to identify all the affected children in the country. Each affected child will be followed for two years, and the outcomes of the children identified by screening will be compared to those of the children diagnosed clinically. It would be difficult to conduct a controlled pilot study of this magnitude in Quebec because of the low annual number of births.

2) In the second scenario, MS/MS implementation would be postponed until the technology transfer can be carried out at the same time for all three diseases, that is, after the necessary validation studies for TT1 have been completed. The transition would thus be simpler. However, this option would involve an additional delay in the provision of MCADD screening and, as a result, postpone its benefits for families.

37. The UK Collaborative Study of Newborn Screening - MCADD is a prospective study whose main objective is to evaluate the clinical and psychological outcomes of children with MCADD and their families by comparing the group of affected children identified by MS/MS-based screening to the group diagnosed clinically [Pollitt, 2006; Oerton et al., 2005; Shortland, 2004]. The study began in March 2004 and is expected to end in 2008, with the possibility of interim results being released beforehand. The investigators planned to conduct neonatal MCADD screening for a 2-year period, to track down all affected newborns not identified through screening using a systematic surveillance throughout the country, and to follow up all affected children for two years, whether they were identified by screening or clinically diagnosed. Interviews with the parents will be analyzed for the purpose of evaluating psychological outcomes. An economic analysis is also planned. Some of the results of this study might be difficult to export to the Quebec context because of the later age at blood sampling in the UK (5 to 8 days of life). Furthermore, the study is using an undervatized MS/MS analytical protocol, unlike the studies on which we based our evaluation of MS/MS performance for the neonatal screening of the three diseases of interest.
with an affected newborn. The planning and preparation, which are discussed below, could nonetheless be initiated beforehand.

3) The third scenario entails immediate MS/MS implementation for PKU and MCADD screening. Thus, there would be no delay for MCADD screening, and investing in the MS/MS technology would be cost-effective from the outset, since the use of this technology is cost-effective as soon as two diseases, including PKU, are screened for. For neonatal TT1 screening, three approaches are possible:

a) Maintaining the current protocols for tyrosine and succinylacetone quantification until the common protocol for all metabolites of interest has been fully validated\(^{38}\). This option therefore involves implementing MS/MS in two phases. The drawback is the temporary use of all of the current analytical methods\(^{39}\) alongside the introduction of MS/MS.

b) Maintaining the current protocols only for succinylacetone quantification and using MS/MS for tyrosine quantification.

c) Quantifying succinylacetone by MS/MS, but with a separate protocol and analytical method (see Chapter 4), using either the same instrument alternately\(^{40}\) or another instrument. In this case, the costs associated with the investment and/or the instrument’s lifespan would potentially be higher.

All three approaches would require coordinating the management of blood samples for TT1 screening on the one hand, and for MCADD and PKU screening on the other. This coordination problem exists in any case for neonatal congenital hypothyroidism screening, since MS/MS does not detect this disease. The literature stresses the importance of taking precautions when implementing MS/MS-based neonatal screening in order not to affect in any way the course or the performance of congenital hypothyroidism screening. Some authors recommend performing MS/MS-based screening and congenital hypothyroidism screening in the same laboratory so as to avoid problems associated with transferring and sharing blood samples between laboratories \cite{Clarke2002}. Common solutions could therefore be considered for neonatal screening of TT1 (approaches a and b) and congenital hypothyroidism. In addition, regardless of the approach chosen, the organization and coordination of the test interpretation need to be closely monitored so as to avoid delays in communicating results. Furthermore, the additional transitional stage, needed if the validation studies of the common analytical protocol quantifying all metabolites are conclusive, should be borne in mind. The potential impact of this additional transition on the course of an already ongoing MS/MS-based screening will need to be discussed and the problems resolved beforehand with the experts.

The different scenarios presented above entail different implications, both in terms of service organization and access to care. The choice between these three options is, of course, based on value judgments, but also depends on more concrete issues. The latter relate to the time required to prepare for technology implementation and to conduct—in Quebec or elsewhere—validation studies of the simplified protocol for TT1 screening, as well as to anticipated difficulties with a phased-in MS/MS implementation. The decision as to the best time to implement MS/MS will involve a trade-off between

\(^{38}\) It should be noted that in this scenario, tyrosine quantification would also be performed by MS/MS for PKU screening in order to calculate the phenylalanine/tyrosine ratio. We did, however, want to keep approaches a and b separate because, depending on the location chosen for implementing MS/MS-based neonatal screening, it may or may not be advisable to continue assaying tyrosine in the same laboratory as succinylacetone for the purpose of a joint interpretation of results.

\(^{39}\) These involve fluorometry for tyrosine quantification and the semi-quantitative method for the succinylacetone assay.

\(^{40}\) Succinylacetone assay could be performed when the MS/MS instrument is not being used for neonatal PKU and MCADD screening, either at night or during weekends or on the day following that of dried blood samples reception, while newly received samples are being prepared for the analysis.
opting to introduce the technology on the basis of data that are scientifically more sound and/or applicable to Quebec and favouring rapid access to services.

Whichever option is chosen, implementing MS/MS must not be done hastily, since other issues—ethical, organizational and economic—need to be resolved beforehand.

- As pointed out in the INSPQ report [Laflamme et al., 2006], the current policy regarding implicit consent for neonatal screening will need to be reviewed. The issue of implicit or explicit consent for neonatal screening is debated in the literature, and the addition of one disease to the screening program will require a reevaluation for the Quebec context in particular. Indeed, the justification for an implicit consent was based on an evaluation of the extent of the benefits and the routine nature of neonatal screening for the three diseases included in the Quebec program.

- Among the values that underpin the health-care system and guide resource allocation, cost-effectiveness is high on the list. The review of the economic literature indicates that MS/MS-based neonatal screening offers an acceptable cost-effectiveness if at least two inborn errors of metabolism, including PKU, are screened for. Many studies are based on rather crude economic modeling and suffer from a lack of detailed cost data, although sensitivity analyses help to compensate for these shortcomings. Even though it would be desirable for this type of modeling to be pursued more thoroughly and to take into account the latest technological developments, these analyses remain tributary of the availability of data applicable to Quebec and on the uncertainty surrounding the clinical data on certain inborn errors of metabolism. It had been agreed that such modeling would not be covered in this report, especially since the INSPQ report had emphasized the difficulty in obtaining accurate cost estimates for current neonatal screening methods. The equivalent annual incremental cost (EAIC) estimated in this report is not modulated according to the different scenarios presented here for MS/MS implementation because of the paucity of field data and the recent publication of one of the discussed protocols. Note that the only available data on the current neonatal screening program are of a completely different order than the EAIC estimates. In fact, the data on the budget allocated to the neonatal screening program [Laflamme et al., 2006] cover costs of a different nature than those considered in this report. A direct comparison of the different operating cost estimates would therefore be risky.

The cost estimates discussed above constitute only one of the necessary elements for a feasibility analysis of MS/MS implementation. A more detailed planning is beyond the scope of this assessment. However, we discuss below several elements that should be considered early on, before a decision in favour of the technology’s implementation is made. These elements concern different stages of MS/MS implementation, namely, the pre-implementation preparations, the operation of the program, and its periodic evaluation.

Several steps will be required before an MS/MS-based neonatal screening program can be launched. Most of them directly concern the laboratory, but steps involving other actors must not be neglected. As regards the laboratory, preparation involves not only laying-out the premises and purchasing the necessary equipment, but also developing and testing all of the processes. The latter include the choice of the metabolites and metabolite ratios for MCADD, the determination of the cut-off values, the evaluation and confirmation of the analytical validity of the protocols, the development of standardized operating procedures (SOPs) and of quality assurance measures, the development of a backup plan to be used during the instrument’s downtime, the establishment of target standards and the planning of the chosen organizational procedures. Recommendations

41. For example, standards will need to be established for the turnaround time for performing the tests and communicating the results.
concerning technical requirements for MS/MS were published by the Centers for Disease Control and Prevention [CDC, 2001]. As for training and information, the provision of training for the laboratory personnel regarding the MS/MS analyses and of training for all health professionals responsible for providing information on MCADD or collecting blood samples, as well as the preparation of educational materials for future parents will all need to be planned. A consensus will have to be reached on the procedures to be followed when non targeted inborn errors of metabolism are detected and on the diagnostic confirmation for MCADD. Lastly, consultations between laboratory directors, clinicians and metabolic disease specialists will be required to determine and standardize the optimal age of sampling and the diagnostic workup for MCADD, and to plan ongoing data collection. Special attention will need to be given to the type of data to be collected, both in the laboratory and at the metabolic disease centres, and to the type of analysis to be performed for the periodic reevaluation of the program. In addition, the information system(s) will have to be adjusted accordingly to facilitate the entry of laboratory and monitoring data, the classification of results, the production of reports and of correspondence relating thereto, the generation of data on program performance and of new knowledge on the screened-for diseases. The budget allocated to the neonatal screening program will have to be revised accordingly, with consideration of the costs associated with long-term data collection in order to permit, particularly for MCADD, an assessment of the benefits of neonatal screening.

Because of the instruments’ complexity and sensitivity, precautions will need to be taken on an ongoing basis once the program starts. These include meticulous instrument maintenance and systematic application of quality assurance mechanisms. Participation in an external quality control program is advisable. Sustained efforts will be necessary to maintain communication and coordination between the various components of the integrated neonatal screening system. The program’s performance will need to be reviewed regularly, and the data on the diseases screened for will need to be reevaluated periodically in light of the monitoring data, the new scientific evidence and the technological advances.

Lastly, it is worth mentioning the frequently raised concerns about the use of the MS/MS technology to perform the full scan of metabolic profiles that detects more than 30 inborn errors of metabolism. Once the MS/MS technology is implemented, there will be increased pressure to expand the neonatal screening program to several other inborn errors of metabolism. This pressure will be exerted by health professionals and industry, as well as by parent associations and the general public, who are increasingly informed through the Internet. The arguments fuelling this pressure include the minimal costs of adding other inborn errors of metabolism once the technology is in place, the advantage of gathering data for research, the benefits for families and the ability to capitalize on what is considered the main advantage of MS/MS, namely, its capacity to analyze several metabolites simultaneously [Pollitt, 2006; Wilcken, 2006; Chace and Kalas, 2005; Roscher and Olgemoller, 2004; Matern, 2002]. Under no circumstances should screening for additional diseases be considered without a prior evaluation of the evidence and criteria that should guide the implementation of population-based screening programs. Finally, several problems discussed in this report, particularly those concerning the provision of information to parents and the availability of an effective network for patient management and follow-up by competent professionals, must necessarily be assessed and solved prior to any expansion of neonatal screening to other inborn errors of metabolism.
In order to examine the relevance of introducing MS/MS in Quebec and expanding neonatal screening to MCADD, a literature review was conducted on the clinical utility of neonatal screening and on MS/MS performance for PKU, TT1 and MCADD screening. The main ethical, social, economic and organizational issues were also investigated. Even though there are gaps in the data, current evidence supports the clinical utility of neonatal screening for most of the patients and families concerned. As for the relevance of implementing MS/MS-based screening in Quebec, the situation differs according to the disease.

- For MCADD, neonatal screening can only be carried out with MS/MS, the performance of which is particularly high for this condition. Knowledge of the entire spectrum of clinical forms is limited, especially for the less severe forms, and the variability in phenotypic expression makes it more difficult to compare the prognosis with and without screening and early management. The benefits of early treatment are convincing, however, for the severe end of the spectrum. Periodic reassessment of MCADD screening benefits, based on ongoing data collection, will therefore be essential.

- For PKU, the literature suggests that MS/MS yields fewer false positives than the current technology, but compared to the results observed in Quebec, this advantage would not be substantial. According to the health economics literature, technology replacement is efficient if screening is carried out for at least two diseases, including PKU. If MS/MS were used for MCADD screening, the technology transfer for PKU would avoid duplication of analytical steps and would probably be less expensive than continuing with the current analytical method alongside MS/MS.

- For TT1, data supporting the efficacy of NTBC therapy are starting to accumulate, and they corroborate the utility of neonatal screening in Quebec. New approaches to TT1 screening based on quantifying both tyrosine and succinylacetone seem promising but need further validation.

Our review confirms the importance of a case-by-case analysis for each inborn error of metabolism. Indeed, the available options depend on the specific characteristics and the state of knowledge for each disease, and the applicability of the technological developments to these diseases [Holtzman, 2003]. Consequently, if expansion to other inborn errors of metabolism is considered, all of the issues raised in this report would have to be reexamined.
For the three diseases of interest, three separate scenarios are therefore proposed for consideration by policy-makers:

1. Conducting a pilot study on the screening for these three diseases over several years.

2. Postponing MS/MS implementation until after validation studies for succinylacetone assays have been completed and a single analytical protocol for neonatal screening of the three diseases can be implemented.

3. Introducing MS/MS-based screening for PKU and MCADD while, either undertaking gradual technology replacement for TT1, or maintaining the current methods until the results of the validation studies are available.

Each scenario has its advantages and disadvantages and entails different organizational issues. The decision as to the best time to implement MS/MS will involve a trade-off between favouring rapid access to services and introducing the technology on the basis of data that are scientifically more sound and/or applicable to Quebec. More thorough data on the recently developed analytical protocols for TT1 screening could be derived from a validation study, and Quebec could be a favourable environment to conduct such a study. As for the advantages of a pilot study, these include collecting epidemiological and genetic data on MCADD and evaluating the costs directly applicable in Quebec. However, it is not likely that such a pilot project would provide, within a reasonable timeframe, the data required for a definite evaluation of the benefits of MCADD screening in terms of long-term prognosis. It will nevertheless be essential to periodically reassess the benefits of neonatal MCADD screening.

Whichever option is chosen, implementing MS/MS must not be done hastily, since a number of issues need to be resolved beforehand. The policy regarding implicit consent for neonatal screening needs to be reviewed, particularly if the decision is taken to add a new disease to the screening program. Indeed, the procedure adopted in Quebec to justify the inclusion of neonatal screening in routine care would pose problems in that event. A more thorough analysis of the feasibility of implementing MS/MS must also be carried out, taking into consideration, amongst other things, the capital and operating costs according to the chosen scenario, and the training costs of health professionals involved throughout the integrated system of neonatal screening. At each stage of a neonatal screening program's implementation, organizational issues must be addressed in order to prospectively optimize practices, by developing standardized protocols and guidelines for instance, and to generate the data needed to monitor the program's performance and to periodically evaluate the pertinence of the choices made.
During the final stage of editing this report, we learned about the publication of an article [Wilcken et al., 2007] evaluating the benefits of MS/MS-based neonatal MCADD screening with regard to the occurrence of death and metabolic crises, and in terms of medical and neuropsychological outcomes. The study was conducted in Australia in a population of approximately 2,500,000 children born between April 1994 and March 2004, of which 810,000 had been screened for MCADD by MS/MS. From this population, the authors followed longitudinally for at least four years 1,995,000 children born between 1994 and 2002, including 460,000 who had been screened. The results of this study support several observations discussed in this report, specifically: the excellent MS/MS performance for MCADD screening, the difference in the genotypic profile between the groups of screened patients and the groups diagnosed clinically, and the importance of reaching a consensus on the diagnosis of the disease. In addition, Wilcken and colleagues [2007] estimated the relative risk (RR) of death or acute metabolic decompensation during the first four years of life on a population scale. To take into account the potential bias due to the differences in the spectrum of the disease between the groups of screened and clinically diagnosed patients, the authors considered several analytical scenarios. Thus, according to the most conservative scenario concerning the prognosis of MCADD cases that escape clinical diagnosis in unscreened children, the benefits of neonatal screening would not be statistically significant (RR=0.44; 95% CI: 0.13-1.45). However, when less conservative scenarios were considered, the risk of death or acute metabolic decompensation was significantly lower in groups of screened newborns than in groups of children diagnosed clinically, with RR ranging from 0.19 (95% CI: 0.06-0.60) to 0.26 (95% CI: 0.08-0.85). In addition, among the children born between 1994 and 2004 and followed for at least two years, the authors recorded far fewer deaths or episodes of severe metabolic decompensation in the cohort of screened newborns (5%) than in the unscreened cohort (55%). The data from this study raise doubts, however, as to the actual proportion of children who would benefit from screening and early management, since four infants died within the first 72 hours of life. In a comment accompanying the publication of Wilcken and colleagues’ article, Grosse and Dezateux [2007] state that screening would prevent one death per 10 children with MCADD, a much more modest figure than that considered in existing cost-effectiveness analyses. These authors stress the need to study the long-term benefits of neonatal screening both for MCADD and for other diseases for which the benefits are still less well known.
Eligibility criteria for study selection

Inclusion criteria

The themes “MS/MS performance for neonatal screening of inborn errors of metabolism” and “screening” for the reviews concerning PKU, TT1 and MCADD:

The following criteria were used to select references relevant to the chapter on MS/MS performance for neonatal screening of inborn errors of metabolism:

- Type of study: Systematic literature reviews or primary research not consisting of case reports
- Subjects: Newborns
- Intervention: MS/MS
- Outcome: Articles providing data on, or allowing calculation of, MS/MS performance criteria for:
  - neonatal screening of groups of inborn errors of metabolism, with the exclusion of articles covered in the British review by Pandor and colleagues [2004];
  - selective neonatal screening for PKU, TT1 or MCADD.

Other themes in the review of the diseases

The inclusion criteria for selecting articles pertaining to the themes of “incidence/prevalence” and “disease treatment” are presented in Tables A-1 and A-2, respectively. As indicated in Table A-1, only articles providing Quebec or Canadian data on the incidence or prevalence of PKU and TT1 were selected. In the case of MCADD, all epidemiological data published after 2000 and North American and European data published before that year were taken into consideration.

Lastly, no specific inclusion criteria were applied for the themes “diagnosis”, “prognosis” and “epidemiology” in the review of the diseases.

| TABLE A-1 |
|------------|----------|----------|----------|
|            | PKU      | TT1      | MCADD    |
| Country    | Quebec, Canada | Quebec, Canada | All | North America, Europe | All |
| Type of study | Reviews, consensus statements or recommendations | Cohort or cross-sectional studies |
| Subjects   | Newborns with the disease |
| Outcome    | Data on, or allowing calculation of, the disease incidence or prevalence |
# Inclusion criteria for articles on the treatment of the diseases

<table>
<thead>
<tr>
<th>Date</th>
<th>PKU</th>
<th>TT1</th>
<th>MCADD</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>≥ 2000</td>
</tr>
</tbody>
</table>

**Type of study**
- Systematic or nonsystematic reviews
  - Primary articles: All clinical trials, cohort studies, and case-control studies
  - Case reports (if n > 2)
  - Systematic reviews: All
  - Narrative reviews: The most systematic

**Subjects**
- Affected adults or newborns

**Intervention**
- Any type of treatment (dietary, pharmacologic, surgical, etc.)

**Outcome**
- Any type

---

**Exclusion criteria**

We excluded from the selected articles studies with any of the following characteristics:

- Language of publication other than English or French.
- Studies strictly involving animals.
- Studies evaluating MS/MS performance for the screening of inborn errors of metabolism other than PKU, TT1 or MCADD.
- Studies evaluating MS/MS performance for urine-based screening of inborn errors of metabolism.
- Studies evaluating neonatal screening of inborn errors of metabolism with a method other than MS/MS.
- Studies on maternal PKU.
- Articles dealing with pediatric liver or kidney diseases, liver transplantation (unless they concerned the evaluation of a treatment for tyrosinemia) or the use of embryonic cells for liver or other transplants.
# Appendix B

## Prevalence of PKU

### TABLE B-1

Quebec or Canadian studies on the prevalence of hyperphenylalaninemia (HPA)

<table>
<thead>
<tr>
<th>REGION REFERENCE</th>
<th>STUDY PERIOD</th>
<th>SOURCE OF DATA</th>
<th>NUMBER OF NEWBORNS SCREENED</th>
<th>NUMBER OF CASES OF PKU DIAGNOSED</th>
<th>PREVALENCE OF CLASSICAL AND ATYPICAL PKU (CASES PER 100,000)</th>
<th>NUMBER OF CASES OF NON-PKU HPA DIAGNOSED</th>
<th>PREVALENCE OF NON-PKU HPA (CASES PER 100,000)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Canada Donlon et al., 2004*</td>
<td>1980-1984</td>
<td>8 provincial neonatal screening programs</td>
<td>2,000,000</td>
<td>91</td>
<td>1:21,966 (4.5)</td>
<td>77</td>
<td>1:25,932 (3.9)</td>
</tr>
<tr>
<td>Quebec Laberge et al., 2005†</td>
<td>1969-1987</td>
<td>Quebec neonatal screening program</td>
<td>1,535,904</td>
<td>60</td>
<td>1:25,598 (3.9)</td>
<td>61</td>
<td>1:25,178 (4.0)</td>
</tr>
<tr>
<td>Quebec Laflamme et al., 2006</td>
<td>1969-2004</td>
<td>Quebec neonatal screening program</td>
<td>Not indicated</td>
<td>Not indicated</td>
<td>1:24,767 (4.0)</td>
<td>Not indicated</td>
<td>1:25,436 (3.9)</td>
</tr>
<tr>
<td>Ontario ACIEM,1976</td>
<td>1965-1975</td>
<td>Ontario neonatal screening program</td>
<td>1,297,100</td>
<td>89</td>
<td>1:14,574 (6.9)</td>
<td>30</td>
<td>1:43,237 (2.3)</td>
</tr>
<tr>
<td>British Columbia Applegarth et al., 2000</td>
<td>1969-1996</td>
<td>Database of the laboratory performing neonatal screening tests</td>
<td>1,142,912</td>
<td>86</td>
<td>1:13,290 (7.5)</td>
<td>26</td>
<td>1:43,958 (2.3)</td>
</tr>
</tbody>
</table>

† Original reference: Grenier et al., 1988.
## Prevalence of TT1

### TABLE C-1

<table>
<thead>
<tr>
<th>STUDY</th>
<th>LOCATION</th>
<th>STUDY PERIOD</th>
<th>PATIENT SELECTION</th>
<th>DISEASE CONFIRMATION</th>
<th>POPULATION SCREENED</th>
<th>NUMBER OF CASES DIAGNOSED</th>
<th>PREVALENCE (CASES PER 100,000)</th>
</tr>
</thead>
<tbody>
<tr>
<td>De Braekeleer and Larochelle, 1990</td>
<td>Quebec, Canada*</td>
<td>October 1970-December 1988</td>
<td>All newborns in Quebec</td>
<td>Yes</td>
<td>1,633,366</td>
<td>98</td>
<td>1:16,667 (6.01)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1967-1971</td>
<td>All newborns in the Saguenay-Lac-Saint-Jean region</td>
<td></td>
<td>21,880</td>
<td>21</td>
<td>1:1042</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1972-1976</td>
<td></td>
<td></td>
<td>22,985</td>
<td>15</td>
<td>1:1532</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1977-1981</td>
<td></td>
<td></td>
<td>28,133</td>
<td>16</td>
<td>1:1758</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1982-1986</td>
<td></td>
<td></td>
<td>24,004</td>
<td>13</td>
<td>1:1846 (54.2)</td>
</tr>
<tr>
<td>Bergeron et al., 1974†</td>
<td>Quebec, Canada*</td>
<td>October 1970-September 1972</td>
<td>82% of newborns in Quebec</td>
<td>Yes</td>
<td>168,727</td>
<td>14</td>
<td>1:12,052 (8.3)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>94% of newborns in the Saguenay-Lac-Saint-Jean region</td>
<td></td>
<td>7522</td>
<td>11</td>
<td>1:684 (146.7)</td>
</tr>
<tr>
<td>Applegarth et al., 2000</td>
<td>British Columbia, Canada‡</td>
<td>1969-1996</td>
<td>All newborns in British Columbia</td>
<td>Yes</td>
<td>1,142,912</td>
<td>8</td>
<td>1:142,864 (0.7)</td>
</tr>
</tbody>
</table>

* These data were collected as part of the Quebec’s neonatal blood screening program for genetic diseases.
† The data presented in this study may be included in De Braekeleer and Larochelle’s study [1990].
‡ The children with inborn errors of metabolism were identified from the registry of the Biochemical Diseases Laboratory, Children’s Hospital, Vancouver.
### Prevalence of MCADD and of the A985G mutation

**TABLE D-1**

<table>
<thead>
<tr>
<th>Study</th>
<th>Origin of the Study Population</th>
<th>Study Design</th>
<th>Study Period</th>
<th>Patient Selection</th>
<th>Population Screened</th>
<th>Number of Cases Diagnosed</th>
<th>Prevalence (Cases per 100,000)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Andresen et al., 2001</td>
<td>Pennsylvania, Ohio, New Jersey, Illinois, Florida, North Carolina</td>
<td>United States Prospective cohort study</td>
<td>December 1, 1992-January 31, 2001</td>
<td>All infants born during this 8-year period</td>
<td>930,078</td>
<td>62</td>
<td>1:15,001 (6.7)</td>
</tr>
<tr>
<td>Zytkovicz et al., 2001</td>
<td>Massachusetts, Maine, New Hampshire, Vermont, Rhode Island</td>
<td>United States Prospective cohort study</td>
<td>1999-2001</td>
<td>164,000 infants born during this period in Massachusetts and 20,000 infants born in Maine</td>
<td>184,000</td>
<td>10</td>
<td>1:18,400 (5.4)</td>
</tr>
<tr>
<td>Chace et al., 2002</td>
<td>Pennsylvania, North Carolina</td>
<td>United States Cohort study</td>
<td>Not indicated</td>
<td>Not indicated</td>
<td>1,019,602</td>
<td>65</td>
<td>1:15,686 (6.4)</td>
</tr>
<tr>
<td>Insinga et al., 2002</td>
<td>Wisconsin</td>
<td>United States Prospective cohort study</td>
<td>April 2000-May 2002</td>
<td>All infants born during this period</td>
<td>155,500</td>
<td>7</td>
<td>1:22 214 (4.5)</td>
</tr>
<tr>
<td>Frazier et al., 2006</td>
<td>North Carolina</td>
<td>United States Prospective cohort study</td>
<td>July 28, 1997-July 28, 2005</td>
<td>All infants born during this period</td>
<td>944,078</td>
<td>73</td>
<td>1:12,933 (7.7)</td>
</tr>
<tr>
<td>STUDY</td>
<td>ORIGIN OF THE STUDY POPULATION</td>
<td>STUDY DESIGN</td>
<td>STUDY PERIOD</td>
<td>PATIENT SELECTION</td>
<td>POPULATION SCREENED</td>
<td>NUMBER OF CASES DIAGNOSED</td>
<td>PREVALENCE (CASES PER 100,000)</td>
</tr>
<tr>
<td>-----------------------</td>
<td>---------------------------------</td>
<td>------------------</td>
<td>-------------------------------</td>
<td>------------------------------------------------------------------------------------</td>
<td>---------------------</td>
<td>---------------------------</td>
<td>--------------------------------</td>
</tr>
<tr>
<td>Feuchbaum et al., 2006b</td>
<td>California, United States</td>
<td>Prospective cohort study</td>
<td>January 2002-June 2003</td>
<td>Screening was offered to 52% of infants born in California during this period, but only 47% of families agreed to screening.</td>
<td>353,894</td>
<td>13</td>
<td>1:27,223 (3.7)</td>
</tr>
<tr>
<td>Wilcken et al., 2003</td>
<td>New South Wales, Australian Capital Territory</td>
<td>Prospective cohort study</td>
<td>April 1998-March 2002</td>
<td>&gt; 99% of the infants born during this period. This study included a sample of 275,653 newborns previously included in the study by Carpenter and colleagues [2001].</td>
<td>362,000</td>
<td>17</td>
<td>1:21,294 (4.7)</td>
</tr>
<tr>
<td>Hoffman et al., 2004</td>
<td>Bavaria, Baden-Würtemberg</td>
<td>Prospective cohort study</td>
<td>January 1999-December 2000</td>
<td>All infants born during this period</td>
<td>382,247</td>
<td>29</td>
<td>1:13,181 (7.6)</td>
</tr>
<tr>
<td>Maier et al., 2005</td>
<td>Bavaria</td>
<td>Prospective cohort study</td>
<td>January 1999-June 2003</td>
<td>All infants born during this period</td>
<td>524,287</td>
<td>62</td>
<td>1:8,456 (11.8)</td>
</tr>
<tr>
<td>Sander et al., 2001</td>
<td>Lower Saxony</td>
<td>Prospective cohort study</td>
<td>1999-2000</td>
<td>Samples from 97% of infants born during this period</td>
<td>127,598</td>
<td>26</td>
<td>1:4,908 (20.3)</td>
</tr>
<tr>
<td>Schulze et al., 2003a</td>
<td>Baden-Würtemberg</td>
<td>Prospective cohort study</td>
<td>April 1998-September 2001</td>
<td>Samples from infants born during this period</td>
<td>250,000</td>
<td>16</td>
<td>1:15,600 (6.4)</td>
</tr>
<tr>
<td>Pourfarzam et al., 2001</td>
<td>Not indicated</td>
<td>Prospective cohort study</td>
<td>January 1, 1991-July 20, 1993</td>
<td>Samples from infants born during this period</td>
<td>100,600</td>
<td>8</td>
<td>1:12,575 (8.0)</td>
</tr>
<tr>
<td>Shigematsu et al., 2002</td>
<td>Not indicated</td>
<td>Prospective cohort study</td>
<td>April 1997-July 2001</td>
<td>Samples from infants born during this period</td>
<td>102,200</td>
<td>2</td>
<td>1:51,100 (1.96)</td>
</tr>
<tr>
<td>Yoon et al., 2005</td>
<td>Not indicated</td>
<td>Prospective cohort study</td>
<td>April 2001-March 2004</td>
<td>Samples from infants born during this period</td>
<td>79,179</td>
<td>0</td>
<td>0 (0)</td>
</tr>
<tr>
<td>STUDY</td>
<td>ORIGIN OF THE STUDY POPULATION</td>
<td>STUDY DESIGN</td>
<td>STUDY PERIOD</td>
<td>NUMBER OF CASES OF MCADD DETECTED</td>
<td>NUMBER OF HOMOZYGOTES</td>
<td>PROPORTION OF HOMOZYGOTES (%)</td>
<td></td>
</tr>
<tr>
<td>---------------------</td>
<td>----------------------------------------------------------</td>
<td>------------------------</td>
<td>-------------------------</td>
<td>----------------------------------</td>
<td>-----------------------</td>
<td>-------------------------------</td>
<td></td>
</tr>
<tr>
<td>Zytkovicz et al., 2001</td>
<td>Massachusetts, Maine, New Hampshire, Vermont and Rhode Island</td>
<td>United States</td>
<td>Prospective cohort study</td>
<td>1999-2001</td>
<td>10</td>
<td>4</td>
<td>40.0</td>
</tr>
<tr>
<td>Carpenter et al., 2001</td>
<td>New South Wales, Australian Capital Territory</td>
<td>Australia</td>
<td>Prospective cohort study</td>
<td>April 1998-March 2001</td>
<td>11</td>
<td>4</td>
<td>36.4</td>
</tr>
<tr>
<td>Zschocke et al., 2001</td>
<td>Not indicated</td>
<td>Germany</td>
<td>Not indicated</td>
<td>Not indicated</td>
<td>14</td>
<td>10</td>
<td>71.4</td>
</tr>
<tr>
<td>Maier et al., 2005</td>
<td>Bavaria</td>
<td>Germany</td>
<td>Prospective cohort study</td>
<td>January 1999-June 2003</td>
<td>57*</td>
<td>27</td>
<td>47.4</td>
</tr>
</tbody>
</table>

* This study identified 62 newborns with MCADD, but molecular analysis could not be performed in five of them.
<table>
<thead>
<tr>
<th>STUDY</th>
<th>ORIGIN OF THE STUDY POPULATION</th>
<th>REGION(S)</th>
<th>COUNTRY</th>
<th>POPULATION SCREENED</th>
<th>NUMBER OF HETEROZYGOTES</th>
<th>PROPORTION OF A985G MUTATION CARRIERS IN THE POPULATION</th>
<th>ESTIMATED PREVALENCE OF CASES OF MCADD WITH HOMOZYGOUS A985G MUTATION</th>
</tr>
</thead>
<tbody>
<tr>
<td>Matsubara et al., 1991</td>
<td>Birmingham, Melbourne, Houston, Sendai</td>
<td>England Australia, United States, Japan</td>
<td>479, 353, 536, 500</td>
<td>12, 5, 5, 0</td>
<td>1:40, 1:71, 1:107, –</td>
<td>1:6400, 1:20,164, 1:45,796, –</td>
<td></td>
</tr>
<tr>
<td>Levin et al., 1992</td>
<td>St. Petersburg</td>
<td>Russia</td>
<td>413</td>
<td>5</td>
<td>1:83</td>
<td>1:27,556</td>
<td></td>
</tr>
<tr>
<td>Dundar et al., 1993</td>
<td>Not indicated</td>
<td>Scotland</td>
<td>552</td>
<td>2</td>
<td>1:276</td>
<td>1:304,704</td>
<td></td>
</tr>
<tr>
<td>Gregersen et al., 1993</td>
<td>Not indicated</td>
<td>Italy</td>
<td>997</td>
<td>3</td>
<td>1:333</td>
<td>1:443,556</td>
<td></td>
</tr>
<tr>
<td>Thompson et al., 1995</td>
<td>Manitoba</td>
<td>Canada</td>
<td>2000</td>
<td>13</td>
<td>1:154</td>
<td>1:94,864</td>
<td></td>
</tr>
<tr>
<td>Conne et al., 1995</td>
<td>Not indicated</td>
<td>Switzerland</td>
<td>1142</td>
<td>22</td>
<td>1:52</td>
<td>1:10,816</td>
<td></td>
</tr>
<tr>
<td>Santer et al., 1995</td>
<td>Schleswig-Holstein (most northern state)</td>
<td>Germany</td>
<td>1000</td>
<td>9</td>
<td>1:111</td>
<td>1:49,284</td>
<td></td>
</tr>
<tr>
<td>Kozak et al., 1999</td>
<td>Moravia</td>
<td>Czech Republic (southeast)</td>
<td>2826</td>
<td>14</td>
<td>1:202</td>
<td>1:163,216</td>
<td></td>
</tr>
<tr>
<td>De Vries et al., 1996</td>
<td>Not indicated</td>
<td>Holland</td>
<td>6195</td>
<td>99</td>
<td>1:63</td>
<td>1:14,000</td>
<td></td>
</tr>
<tr>
<td>Lecoq et al., 1996</td>
<td>Normandy</td>
<td>France</td>
<td>2000</td>
<td>17</td>
<td>1:118</td>
<td>1:55,700</td>
<td></td>
</tr>
<tr>
<td>Seddon et al., 1997</td>
<td>West Midlands, Trent</td>
<td>England</td>
<td>10,171</td>
<td>158</td>
<td>1:64</td>
<td>1:16,384</td>
<td></td>
</tr>
<tr>
<td>Johansson et al., 1999</td>
<td>Not indicated</td>
<td>Sweden</td>
<td>1015</td>
<td>8</td>
<td>1:127</td>
<td>1:64,516</td>
<td></td>
</tr>
</tbody>
</table>
APPENDIX E

MS/MS Analysis

The MS/MS DEVICE

A tandem mass spectrometer consists of four main components: an ion source, a mass analyzer, an electron detector, and a computer system connected to the detector [Banta-Wright and Steiner, 2004; Cheillan et al., 2004; Rinaldo et al., 2004; Dooley, 2003; Carpenter and Wiley, 2002; Clarke, 2002]. Figure E1 displays the different components of the MS/MS device, particularly the analyzer components.

Adapted from Banta-Wright and Steiner, 2004; Fearing and Levy, 2003; and Clarke, 2002.
The steps of MS/MS analysis

Ionization is a key step, since only ionized molecules can be identified and quantified by mass spectrometry. Ionization results in the molecules of interest acquiring a charge and enables the ions to enter a gaseous phase, another necessary condition for their analysis by MS/MS [McCandless, 2004; Chace et al., 2002]. Ionization can be performed in a positive or negative mode, the choice depending on the chemical structure of the targeted analytes [Cheillan et al., 2004; Adeli, 2003]. There are several types of ionization sources [Cheillan et al., 2004; Pandor et al., 2004]. According to many authors, electrospray ionization (ESI) is the type of choice for the combined analysis of amino acids and acylcarnitines for neonatal screening of inborn errors of metabolism [Cheillan et al., 2004; Dooley, 2003; Carpenter and Wiley, 2002; Chace et al., 2002; Clarke, 2002; Rashed et al., 1995]. Developed around 1995, ESI is used mainly for the analysis of small, charged, nonvolatile molecules in a liquid phase. Its advantages include automated sample injection and the use of a low energy ionization process (soft ionization), which results in little or no fragmentation in the molecules [Banta-Wright and Steiner, 2004; Rashed et al., 1995].

Once produced, the charged ions undergo the effects of an electromagnetic field that will move them to the mass analyzer. In general, the analyzer consists of a series of three chambers (Q1, Q2 and Q3, respectively\(^{42}\)) including two mass spectrometers (hence the designation “tandem”) separated by a collision cell. The advantage of this instrument is inherent to its capacity to quantify parent/precursor ions of interest (intact ionized molecules of the initial mixture) in Q1, to fragment these molecules in Q2\(^{43}\) and to analyse the product/daughter ions in Q3. For each molecule of interest from the initial mixture, the information from Q1 and Q3 is matched in “parent ion/daughter ion” pairs identified by their respective m/z ratios. This information is then captured by a detector and transmitted to a computer system [Banta-Wright and Steiner, 2004; Cheillan et al., 2004; Rinaldo et al., 2004; Chace et al., 2003, 2002; Dooley, 2003; Fearing and Levy, 2003; Carpenter and Wiley, 2002; Clarke, 2002]. Several computer programs allowing automated manipulation of the data and interpretation of results have been developed. Depending on the program, the results may be presented in the form of mass spectrum graphs and/or a spreadsheet. In addition, mathematical formulas have been developed for calculating metabolite ratios, such as the ratio of a metabolite to its internal standard or the ratio of different metabolites\(^{44}\). Lastly, the cut-off values for each metabolite can be automatically incorporated into the computerized systems.

The modes of MS/MS analysis

Scan modes

- **Product ion scan**: This mode is used to detect all product ions resulting from a given precursor ion corresponding to a molecule of interest. To that end, Q1 is programmed to allow only a single precursor ion with a specific m/z ratio to enter Q2, and Q3 is adjusted to detect all product ions derived from this precursor ion\(^{45}\). This is the most suitable method for identifying an unknown substance in a complex mixture and it is used mainly for the validation of analytical methods.

---

\(^{42}\) These chambers are called quadrupoles, hence the Q1, Q2 and Q3 designation.

\(^{43}\) This fragmentation is achieved by a process called collision-induced dissociation (CID).

\(^{44}\) This option is used to characterize certain inborn errors of metabolism, such as PKU for which the phenylalanine/tyrosine ratio is used.

\(^{45}\) Product ions ranging from the lowest m/z ratio to the m/z ratio of the specific precursor ion being analyzed are thus detected in Q3.
- **Precursor ion scan**: This mode is used to detect all the precursor ions in Q1 that produce a given specific product ion. It is utilized to specifically analyze a family of molecules with a common structural part and which therefore yield a common product ion. In this case, Q1 is programmed to allow all the precursor ions from the sample to enter Q2, while Q3 is programmed to allow only product ions with a specific m/z ratio to reach the detector. A spectrum of all the precursor ions producing this daughter ion is thus obtained. In the context of neonatal screening of inborn errors of metabolism, this mode is used to analyze acylcarnitines.

- **Neutral loss scan**: This mode is used to detect all precursor ions sharing a common neutral fragment lost after fragmentation. Hence, Q1 and Q3 are programmed in relation to a constant difference in mass, which is the difference between the mass of the precursor ion and that of the product ion and is equal to the mass of the neutral fragment\(^{46}\). In the context of neonatal screening of inborn errors of metabolism, this scan mode is used for the semi-quantitative analysis of amino acids.

**Single reaction monitoring**

Unlike the three scan modes discussed above, which involve the detection of a range of m/z ratios (in Q1 or Q3), the SRM (single-reaction monitoring) mode targets a single value of m/z ratio in each of the mass spectrometers. The SRM mode is thus used to identify the specific “parent ion/daughter ion” pair which characterizes the molecule of interest. Q1 is programmed to allow only the specific parent ion to travel to Q2, and Q3 is programmed to detect only the corresponding product ion\(^{47}\). A unique fragmentation process for each pair must therefore be carried out, hence the expression “single reaction monitoring”. However, since MS/MS can be used to simultaneously analyze several pairs of this type, the term most often used for this analytical mode is multiple reaction monitoring (MRM), which refers to all of the SRMs performed concurrently. This mode is mainly used to simultaneously quantify a very large number of compounds with different chemical structures. To this end, several adjustment cycles of the magnetic fields are computer-programmed to succeed each-other within a few seconds. Each cycle corresponds to a specific m/z ratio. In the context of neonatal screening of inborn errors of metabolism, the MRM mode is used to selectively analyze amino acids and acylcarnitines that are not adequately detected by neutral loss scans (amino acids) or precursor ion scans (acylcarnitines).\(^{48}\) Furthermore, with the MRM mode the analysis can be restricted to certain metabolites of interest while avoiding the detection of diseases whose natural course and/or treatment are less well known.

\(^{46}\) The neutral fragment cannot be detected directly by the mass spectrometer because it is not ionized.

\(^{47}\) The chosen fragment ion is usually the most abundant ion derived from this precursor ion.

\(^{48}\) This situation arises, for example, for the quantification of hexose-1-phosphate for galactosemia screening, the quantification of steroids for the second-tier screening test for congenital adrenal hyperplasia, and the quantification of succinylacetone for neonatal TT1 screening.
Main analytical steps of the amino acid and acylcarnitine profile

TABLE F-1

Main steps in the MS/MS-based analysis of the common amino acid and acylcarnitine profile used in neonatal screening

Sample preparation

1. Extraction of amino acids and acylcarnitines
   ▪ Punching of the Guthrie cards and loading onto multiwell plates
   ▪ Elution in a solution containing methanol and the internal standards
   ▪ Incubation and shaking during extraction
   ▪ Transfer of the solution containing the extracts to a second multiwell plate
   ▪ Elimination of the excess methanol through evaporation

2. Derivatization
   ▪ Addition of butanolic acid to the extracts
   ▪ Incubation at a temperature that promotes the conversion of the metabolites to butyl esters
   ▪ Elimination of the excess butanolic acid through evaporation
   ▪ Reconstitution in solution

MS/MS analysis

1. Automated aliquot sampling and injection into the instrument at a programmed flow

2. Electrospray ionization

3. Entry into the analyzer programmed for a sequential performance of three scan modes
   ▪ A precursor ion scan with an m/z ratio of 85 Da for acylcarnitine analysis
   ▪ A neutral loss of 102 Da scan for the analysis of acid and neutral amino acids
   ▪ A neutral loss of 119 Da scan for the analysis of basic amino acids

4. Detection and production of results
   ▪ Detection of the signal corresponding to the selected range of m/z ratios
   ▪ Quantification by reference to the internal standards
   ▪ Computer conversion of the signals into a mass spectrum
   ▪ Comparison of the metabolite concentrations and ratios to the pre-established cut-off values
Advantages and disadvantages of MS/MS

**Advantages**

Several advantages are attributed to the MS/MS technology [Cheillan et al., 2004; Fearing and Levy, 2003; Carpenter and Wiley, 2002; Clarke, 2002]. The advantages related to the range of analyses that can be performed with this technology and its potential for wide-scale use are summarized below, whereas the data on MS/MS performance for neonatal screening were presented in Chapter 5.

- The MS/MS technology allows for neonatal screening of a large number of inborn errors of metabolism in a single analytical step. In addition, for some of these errors of metabolism (e.g., MCADD), no other screening test is available.
- MS/MS is versatile in that a wide variety of metabolites can be detected and the analysis can be programmed either to screen for a group of diseases or to selectively target certain diseases.
- The analysis can be performed on blood and urine samples, and it has been adapted for samples of dried blood, a very convenient substrate for neonatal screening (only a small quantity of blood needs to be obtained, and samples can be mailed). If blood for dried blood samples for MS/MS analysis is more time-consuming than the MS/MS analysis per se, which is completely automated.
- Very small quantities of metabolites can be detected, separated and identified.
- MS/MS analysis does not require a prior chromatographic separation of the metabolites because of the concurrent use of two mass spectrometers.
- The total time, from ionization to results generation, including the commutation between the amino acid and the acylcarnitine analysis modes, is approximately two to three minutes per sample. For practical reasons, the samples are prepared and analyzed on 96-well microplates. For each microplate, three hours should be allowed for the sample preparation and five hours for the analysis.
- The analysis is automated and permits a high throughput of approximately 600 samples per 24 hours.

**Disadvantages**

Despite the undeniable contribution of MS/MS, several drawbacks of this technology have been reported [Chace et al., 2005; Cheillan et al., 2004; Fearing and Levy, 2003; Carpenter and Wiley, 2002; Clarke, 2002]. As indicated below, these primarily concern the technical requirements, which are rather heavy, with the need to take precautions at every step, from blood sampling to results interpretation. Certain drawbacks are due to the specific application of MS/MS to neonatal screening, either because of the type of substrate or because of the time of sampling. Lastly, there are issues related to many interpretation problems involving the metabolites of interest and contaminants and to differential diagnosis of diseases.

- The different steps for preparing dried blood samples for MS/MS analysis are more time-consuming than the MS/MS analysis per se, which is completely automated.
- Like other screening techniques, MS/MS analysis depends on the volume of the blood samples and on the conditions of obtaining, transporting and storing them. The necessary quantity of blood for dried blood samples varies according to the hematocrit, the diameter of the punch, the degree of saturation and of hemolysis, and the quality of the filter paper, as well as to the environmental (humidity and...
temperature), transport and storage conditions for the Guthrie Cards\textsuperscript{49} [Holub et al., 2006; Lindner et al., 2006; Al-Dirbashi et al., 2005; Nagy et al., 2003; Santer et al., 2003].

- Because of the complex makeup of the blood matrix, ionization may be hindered by the interaction between the metabolites of interest and ions such as sodium and chloride, or by competition between these metabolites and other molecules. The consequences can be minimized by performing the extraction step with great care and using internal standards.

- The derivatization step has certain drawbacks. This step adds about 45 minutes to the sample preparation time and, like any other analytical step, increases the risk of errors. These errors may be associated with problems such as instability of the derivatized metabolites or interference with the reagents. In addition, two major problems associated with asparagine and glutamine quantification and free carnitine quantification have been attributed to derivatization (Appendix G of the technical report [Makni et al., 2007] provides additional information on these problems).

- The concentration of certain metabolites may be too low for their detection during the neonatal period, even in the presence of an inborn error of metabolism, either because of insufficient protein ingestion or because of the age at sampling. Indeed, the amino acid concentration tends to increase with age, while the concentration of acylcarnitines, particularly the long-chain acylcarnitines, tends to decrease [Cavedon et al., 2005].

- Certain drugs and other substances, such as valproic acid, pivalic acid and benzoic acid, produce acylcarnitines that are detected by MS/MS and whose profile can overlap that of the metabolites of interest. For example, valproic acid, an anticonvulsant, yields a C8 derivative that can be confused with the metabolic profile of MCADD. The differentiation can be made on the basis of the C8/C10 ratio, which is high only if the newborn is actually affected with MCADD. Intake of vitamins or nutritional supplements, as well as parenteral hyperalimentation, can also distort MS/MS results and lead to artificially high amino acid levels, for example.

- Acylcarnitines are not generally specific to a given disease. For this reason, neonatal screening of inborn errors of organic acid and fatty acid metabolism is based on the analysis of several metabolites. Indeed, the metabolic pathways of fatty acid oxidation overlap, and the deficiency in a given enzyme can manifest by an elevated concentration of more than one type of acylcarnitine. For example\textsuperscript{50}, C8 is generally elevated in MCADD but can also be elevated in multiple acyl-CoA dehydrogenase (MAD) deficiency, also referred to as glutaric aciduria type II (GAIID), and in medium/short-chain L-3-hydroxyacyl-CoA dehydrogenase deficiency (M/SCHADD). The differential diagnosis is made on the basis of the C8/C10 ratio, which is elevated in MCADD. Diagnostic confirmation necessarily requires additional tests.

- MS/MS analysis cannot be used to quantify isobaric compounds, such as leucine, isoleucine and hydroxyproline, separately. However, for screening purposes, a semi-quantitative assay is acceptable. Precise quantification of these isobaric compounds can be performed at the time of diagnostic confirmation [Casetta et al., 2000]\textsuperscript{51}.

\begin{footnotesize}
\textsuperscript{49} For example, particles released from the filter paper and ruptured red blood cells can result in a major loss of detection signal in ESI-MS/MS.

\textsuperscript{50} Other examples concern long-chain hydroxyacyl-CoA dehydrogenase (LCHAD) deficiency and trifunctional protein deficiency (C16OH); methylmalonic acidemia, propionic acidemia, biotin deficiencies and vitamin B\textsubscript{12} deficiency (C3), isovaleric acidemia and methylbutyryl-CoA dehydrogenase deficiency (CS).

\textsuperscript{51} Such problems are encountered in maple syrup urine disease, short-chain acyl-CoA dehydrogenase (SCAD) deficiency and isobutyryl-CoA dehydrogenase deficiency (C4 and isobutyrylcarnitine). A detailed analysis of the fragment ratios can sometimes be a solution [Carpenter and Wiley, 2002]. Otherwise, the LC-MS/MS technology, which includes an additional, yet short, chromatographic separation step of about 4 minutes, can be used. In addition, the MRM mode must be used to individually quantify these isobaric amino acids [Nagy et al., 2003].
\end{footnotesize}
Selection of primary articles on MS/MS performance

Figure H-1

n = 306 articles selected through the search strategies on:
- a) MS/MS performance
- b) Screening for inborn errors of metabolism

n = 293 rejected on the basis of titles and abstracts

n = 13 selected

n = 11 rejected after reading the article, including:
- 5 containing no performance data
- 6 case reports

n = 2 selected
Figure H-2

$n = 453$ articles selected through the search strategies on:

a) MS/MS performance

b) Screening for inborn errors of metabolism

$n = 185$ rejected on the basis of titles and abstracts

$n = 191$ selected for other sections of the report

$n = 77$ selected

$n = 67$ rejected after reading the article, including:

- 1 reviewed by Pandor and colleagues
- 10 containing no performance data
- 36 case reports
- 1 concerning a population already included in another study
- 7 pertaining to other inborn errors of metabolism
- 11 selected for other sections of the report

$n = 11$ selected
### Characteristics of primary studies on MS/MS performance for neonatal screening of inborn errors of metabolism

<table>
<thead>
<tr>
<th>IEM AUTHORS, YEAR</th>
<th>COUNTRY</th>
<th>STUDY PERIOD</th>
<th>POPULATION*</th>
<th>STUDY DESIGN</th>
</tr>
</thead>
<tbody>
<tr>
<td>Zytkovicz et al., 2001</td>
<td>New England, United States</td>
<td>1999-2001</td>
<td>Infants born in Massachusetts, Maine, New Hampshire, Vermont and Rhode Island</td>
<td>Prospective cohort</td>
</tr>
<tr>
<td>Shigematsu et al., 2002</td>
<td>Japan</td>
<td>Not indicated</td>
<td>Not indicated</td>
<td>Not indicated</td>
</tr>
<tr>
<td>Schulze et al., 2003a</td>
<td>Germany</td>
<td>1998-2001</td>
<td>Infants born in Baden-Württemberg between April 1998 and September 2001</td>
<td>Prospective cohort</td>
</tr>
<tr>
<td>Feuchtbaum et al., 2006b</td>
<td>California, United States</td>
<td>2002-2003</td>
<td>Genetic Disease Branch, California Department of Health Services, Genetic Disease Laboratory</td>
<td>Prospective cohort</td>
</tr>
<tr>
<td>Frazier et al., 2006</td>
<td>North Carolina, United States</td>
<td>2003-2004</td>
<td>North Carolina’s Tandem Mass Spectrometry Newborn Screening Program</td>
<td>Prospective cohort</td>
</tr>
</tbody>
</table>

*Only the study populations on which the MS/MS performance criteria calculations were based are described.

**Abbreviations:** MS/MS: tandem mass spectrometry; IEM: inborn error of metabolism; DBS: dried blood samples; PKU: phenylketonuria; TT1: tyrosinemia type 1; MCADD: medium-chain acyl-CoA dehydrogenase deficiency.

**Although this was an 8-year study (1997-2005), the article includes MS/MS performance data for neonatal screening of inborn errors of metabolism relative to one year only (2003-2004).
<table>
<thead>
<tr>
<th>IEM</th>
<th>AUTHORS, YEAR</th>
<th>COUNTRY</th>
<th>STUDY PERIOD</th>
<th>NEONATAL SCREENING PROGRAM</th>
<th>POPULATION*</th>
<th>STUDY DESIGN</th>
</tr>
</thead>
<tbody>
<tr>
<td>PKU</td>
<td>Chace et al., 1998</td>
<td>California, United States</td>
<td>Not indicated</td>
<td>California Newborn Screening Program</td>
<td>Samples stored since 1992-1994 and originally tested by fluorometry within the first 24 hours of life</td>
<td>Retrospective analysis of archived DBS</td>
</tr>
<tr>
<td></td>
<td>Ceglarek et al., 2002</td>
<td>Germany</td>
<td>Not indicated</td>
<td>Not indicated</td>
<td>Not indicated</td>
<td>Retrospective analysis of DBS</td>
</tr>
<tr>
<td>TT1</td>
<td>Sander et al., 2006</td>
<td>Germany</td>
<td>Not indicated</td>
<td>Neonatal screening laboratory, Hannover</td>
<td>Residual DBS previously used in the routine neonatal screening program</td>
<td>Prospective cohort</td>
</tr>
<tr>
<td>MCADD</td>
<td>Chace et al., 1997</td>
<td>Pennsylvania, United States</td>
<td>1992-1997</td>
<td>Supplemental Newborn Screening Program, Neo Gen Screening, Pennsylvania</td>
<td>16,500 DBS obtained from the North Carolina Division of Laboratory Services Newborn Screening Program between 1993-1995</td>
<td>Prospective cohort</td>
</tr>
<tr>
<td></td>
<td>Andresen et al., 2001</td>
<td>United States</td>
<td>1992-2001</td>
<td>Name of program not indicated, United States</td>
<td>Newborns in Pennsylvania, Ohio, New Jersey, Illinois, Florida and North Carolina</td>
<td>Prospective cohort</td>
</tr>
<tr>
<td></td>
<td>Carpenter et al., 2001</td>
<td>Australia</td>
<td>1998-2001</td>
<td>New South Wales Newborn Screening Programme</td>
<td>&gt; 99% of neonates born in New South Wales and the Australian Capital Territory between April 1998 and March 2001</td>
<td>Prospective cohort</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>13 newborns with MCADD born between January 1981 and June 1997</td>
<td>Retrospective analysis of archived DBS</td>
</tr>
<tr>
<td></td>
<td>Pourfarzam et al., 2001</td>
<td>United Kingdom</td>
<td>Not indicated</td>
<td>Northern Region of the National Health Service</td>
<td>Neonates born between January 1, 1991 and July 20, 1993 in the Northern Region of the British National Health Service</td>
<td>Retrospective cohort</td>
</tr>
</tbody>
</table>

Abbreviations: MS/MS: tandem mass spectrometry; IEM: inborn error of metabolism; DBS: dried blood samples; PKU: phenylketonuria, TT1: tyrosinemia type 1; MCADD: medium-chain acyl-CoA dehydrogenase deficiency.

* Only the study populations on which the MS/MS performance criteria calculations were based are described.
**APPENDIX J**

**MS/MS performance for neonatal screening of groups of inborn errors of metabolism**

**TABLE J-1**

<table>
<thead>
<tr>
<th></th>
<th>Zytkovicz et al., 2001*</th>
<th>Shigematsu et al., 2002†</th>
<th>Wilcken et al., 2003‡</th>
<th>Schulze et al., 2003§</th>
<th>Frazier et al., 2006∥</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Scenario</strong></td>
<td>For IEM for which the number of infants tested = 257,000</td>
<td>For IEM for which the number of infants tested = 164,000</td>
<td>Scenario a</td>
<td>Scenario b</td>
<td>Scenario a</td>
</tr>
<tr>
<td><strong>Prevalence (%)</strong></td>
<td>0.008</td>
<td>0.006</td>
<td>0.01</td>
<td>0.03</td>
<td>0.04</td>
</tr>
<tr>
<td><strong>True positives (n)</strong></td>
<td>20</td>
<td>10</td>
<td>11</td>
<td>96</td>
<td>106</td>
</tr>
<tr>
<td><strong>False positives (n)</strong></td>
<td>84</td>
<td>244</td>
<td>581</td>
<td>520</td>
<td>851</td>
</tr>
<tr>
<td><strong>False negatives (n)</strong></td>
<td>?</td>
<td>?</td>
<td>7</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td><strong>True negatives (n)</strong></td>
<td>256,896</td>
<td>163,746</td>
<td>101,606</td>
<td>361,377</td>
<td>249,039</td>
</tr>
<tr>
<td><strong>Sensitivity (%)</strong></td>
<td>?</td>
<td>?</td>
<td>91.66</td>
<td>93.20</td>
<td>96.36</td>
</tr>
<tr>
<td><strong>Positive predictive value (%)</strong></td>
<td>19.23</td>
<td>3.94</td>
<td>2.02</td>
<td>15.58</td>
<td>11.08</td>
</tr>
<tr>
<td><strong>Recall rate (%)</strong></td>
<td>0.04</td>
<td>0.16</td>
<td>0.58</td>
<td>0.17</td>
<td>0.38</td>
</tr>
</tbody>
</table>

* The number of newborns tested varied according to the disease: 257,000 for PKU, maple syrup urine disease and hypermethioninemia; 164,000 for all the other inborn errors of amino acids and acylcarnitines metabolism, except MCADD; and 184,000 for MCADD. For this reason, the results for MCADD are not presented in this table, but only in Section 5.2.2.4 and in Table 3, which provide MS/MS performance results for the selective screening for this disease. For inborn errors of amino acid metabolism, newborns were considered as positives when they had an elevated level of the main metabolite and of the ratio of metabolites that were used as markers for the corresponding disease.

† The article only provides the number of false positives and the recall rate, and it is on the basis of the latter that the total number of positive MS/MS results was estimated at 593 and that the performance criteria were calculated.

‡ 26 subjects for whom a diagnosis was suspected but not formally confirmed were considered among the false positives in scenario (a) and among the true positives in Scenario (b). These included 9 subjects lost to follow-up and 17 with a suspected inborn error of metabolism for which the diagnosis was difficult to establish.

§ Scenario (a) considers as positives (true and false positives) subjects with two tests above the borderline cut-off value or one test above the diagnostic cut-off value; Scenario (b) considers as positives subjects with one test above the diagnostic cut-off value; and Scenario (c) considers as positives subjects with one test above the borderline or diagnostic cut-off value. For this study, the authors report four false negatives but do not indicate the period during which this observation was made, especially in relation to the year 2003, the year when the cut-off values were established and on which the performance results presented in the table are based. For this reason, two values are indicated for the number of false and true negatives and for the sensitivity and the negative predictive value. It should be noted that the authors also report two other false negatives. However, they clearly state that these cases were detected before the final determination of the cut-off values, hence, before 2003. They were therefore excluded from the performance criteria calculation.

∥ Scenario (a): 240 subjects not referred for diagnostic confirmation after the clinical review and 75 cases referred, but for whom a final diagnosis was not established (refusal of screening, lost to follow-up, death, etc.) are considered as false positives; Scenario (b): 240 subjects not referred for diagnostic confirmation after the clinical review and the 75 cases referred, but for whom a final diagnosis was not established (refusal of screening, lost to follow-up, death, etc.) are considered as true positives; Scenario (c): the 75 cases referred for diagnostic confirmation but for whom a final diagnosis was not established (refusal of screening, lost to follow-up, death, etc.) are considered as false positives; and Scenario (d): the 75 cases referred for diagnostic confirmation but for whom a final diagnosis was not established (refusal of screening, lost to follow-up, death, etc.) are considered as true positives.
### Economic aspects: literature review

**TABLE K-1**

<table>
<thead>
<tr>
<th>AUTHORS, COUNTRY</th>
<th>DESCRIPTION OF STUDY AND METHODOLOGY</th>
<th>COST AND EFFICACY DETAILS CONSIDERED</th>
<th>MAIN RESULTS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tran <em>et al.</em>, 2006 (CCOHTA) Canada</td>
<td>Review on the use of MS/MS for neonatal MCADD screening as compared to clinical diagnosis. Two economic studies selected according to the criteria set out in the BMJ 35-item checklist [Venditti <em>et al.</em>, 2003; Insinga <em>et al.</em>, 2002]</td>
<td>Costs (2005 Canadian dollars) concerning a provincial laboratory in Nova Scotia: one MS/MS instrument, reagents, personnel, complications, disabilities and deaths associated with MCADD</td>
<td>The literature and the economic analysis show that MS/MS-based screening consumes more resources but yields better health outcomes than no screening (reduced morbidity and mortality). Most screened MCADD patients were asymptomatic, while those diagnosed clinically showed irreversible damage.</td>
</tr>
<tr>
<td></td>
<td>Cost-effectiveness analysis with decision model (tree) and sensitivity analysis (3 scenarios: base-case, best and worst)</td>
<td>Unit cost of screening: $2.40</td>
<td>ICER = $2,514/QALY. Incremental cost = $389,118 in the best scenario; $596,075 in the worst scenario. ICER = $928/QALY in the best scenario; $11,456/QALY in the worst scenario.</td>
</tr>
<tr>
<td></td>
<td>Health-care system perspective</td>
<td>Discount rate: 3%</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Assumption: birth prevalence of MCADD in Canada = 1:16,000</td>
<td>Modelling based on a cohort of 330,803 newborns</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Analysis based on Nova Scotia experience (8,533 newborns per year)</td>
<td>Anticipated costs of screening versus no screening; anticipated effectiveness of screening versus no screening (number of QALYs gained, life-years gained, number of cases detected before onset of symptoms, number of hospitalizations avoided, number of cases of morbidity and death avoided); incremental cost-effectiveness ratio (ICER) of screening versus no screening</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Timeframe for budget impact analysis: 5 years; modelling: 77 years and 66 years</td>
<td>Threshold = $50,000/QALY (literature) and threshold = $20,000/QALY (Canadian context)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Threshold = $50,000/QALY (literature) and threshold = $20,000/QALY (Canadian context)</td>
<td>26% probability of death in unscreened children with MCADD</td>
<td></td>
</tr>
</tbody>
</table>

Abbreviations: QALY: quality-adjusted life-year; ICER: incremental cost-effectiveness ratio; MCADD: medium-chain acyl-CoA dehydrogenase deficiency; MS/MS: tandem mass spectrometry.
<table>
<thead>
<tr>
<th>AUTHORS, COUNTRY</th>
<th>DESCRIPTION OF STUDY AND METHODOLOGY</th>
<th>COST AND EFFICACY DETAILS CONSIDERED</th>
<th>MAIN RESULTS</th>
</tr>
</thead>
</table>
| Pandor et al., 2006, United Kingdom | ▪ Cost-effectiveness analysis of MS/MS-based PKU and MCADD screening  
▪ Probabilistic modelling of newborn screening from a health-care system perspective.  
▪ Update of a 2004 report | ▪ Estimate of the incremental cost (1997 pounds updated to reflect 2001 prices) and of the number of life-years gained.  
▪ Discount rate: 6% for the costs and 1.5% for life-years gained | ▪ Substituting MS/MS for existing PKU screening technologies increases costs with no increase in health outcomes.  
▪ The addition of MCADD to PKU would generate cost savings of 17,298 £ for each cohort of 100,000 neonates screened.  
▪ A gain in the number of life-years is anticipated (57.3 years) as well. |

| Autilt-Rämö et al., 2005, Finland | ▪ Pilot study evaluating the cost-effectiveness of expanding MS/MS-based neonatal screening to five inborn errors of metabolism (CAH, MCADD, LCHADD, PKU and GAI) Modelling and sensitivity analysis based on published data, health-care registers and expert opinions  
▪ Health-care system perspective  
▪ 40% probability of death for unscreened children with MCADD  
▪ Among survivors, 30% probability of mild impairment and 30% probability of severe neurological disability | ▪ The costs considered (2002 euros) include the cost of setting up a new screening unit; the running costs (materials, time, personnel, laboratory space, equipment, and quality control)  
▪ Efficacy data: disease incidence data (panel of experts), data on the anticipated incidence (reliable for CAH and LCHADD); effects on the quality of life (profile of health-related quality of life using the 16D measure  
▪ Discount rate: 5% | ▪ The annual cost of screening for 56,000 newborns is estimated at €2.5 million (2002 price level), or €45 per newborn, to identify 5 to 10 cases annually.  
▪ The costs per QALY gained vary from €5,500 (min.) to €25,500 (max.), depending on the incidence data and QALY weightings.  
▪ If severe handicaps are taken into account, the maximum cost falls to €18,000 per QALY (range: €3,900 to €18,000). |

Abbreviations: QALY: quality-adjusted life-year; CAH: congenital adrenal hyperplasia; GAI: glutaric acidemia type 1; LCHADD: long-chain 3-hydroxyacyl-CoA dehydrogenase deficiency; MCADD: medium-chain acyl-CoA dehydrogenase deficiency; MS/MS: tandem mass spectrometry; PKU: phenylketonuria.
<table>
<thead>
<tr>
<th>AUTHORS, COUNTRY</th>
<th>DESCRIPTION OF STUDY AND METHODOLOGY</th>
<th>COST AND EFFICACY DETAILS CONSIDERED</th>
<th>MAIN RESULTS</th>
</tr>
</thead>
</table>
| Venditti et al., 2003 United States | ▪ Analysis of the efficiency (cost-effectiveness and cost-utility ratios) of neonatal MCADD screening by MS/MS versus no screening  
▪ Markov model for estimating the discounted incremental cost per life-year saved and per QALY gained, and sensitivity analysis of the key variables (second-order Monte Carlo simulations, 95% confidence intervals)  
▪ Societal perspective  
▪ Time horizons of 20 and 70 years for the estimated health states  
▪ Patients categorized according to their diagnosis  
▪ Probability of neurological disability in survivors: 10% | ▪ Costs (2001 US dollars) and probabilities derived from a retrospective chart review of 32 patients treated over 30 years at the Children’s Hospital of Philadelphia, clinical experience with MCADD patient management, patient and family interviews, cost surveys and published studies  
▪ The costs include the screening and follow-up, the confirmatory tests and carnitine quantification for screen-positives, and care for severely affected patients  
▪ ICER estimates  
▪ Assumption: MS/MS operating costs are already covered by its use for PKU screening  
▪ Discount rate: 3% | ▪ Base-case scenario: cost = $11,000/life-year saved over the first 20 years; ICER = $5,600/QALY versus no screening. Over 70 years: cost = $300/life-year saved; ICER = $100/QALY. Neonatal MCADD screening reduces morbidity and mortality at a lower incremental cost than that accepted for this type of intervention. Over 70 years, all the additional costs of screening would be offset by avoided sequelae.  
▪ Varying the model inputs within the ranges chosen for sensitivity analysis did not affect the efficiency of screening. |
| Medical Advisory Secretariat (MAS), 2002 Canada | Systematic review on the cost-effectiveness of MS/MS-based neonatal screening of inborn errors of metabolism | ▪ No cost estimates made  
▪ Acknowledgement that the initial costs are very high (purchase of equipment; information technology support and experienced personnel for determining the appropriate cut-off values for each test) | ▪ Expanding an MS/MS-based screening program beyond PKU and MCADD would probably be cost-effective |

Abbreviations: QALY: quality-adjusted life-year; ICER: incremental cost-effectiveness ratio; MCADD: medium-chain acyl-CoA dehydrogenase deficiency; MS/MS: tandem mass spectrometry; PKU: phenylketonuria.
<table>
<thead>
<tr>
<th>AUTHORS, COUNTRY</th>
<th>DESCRIPTION OF STUDY AND METHODOLOGY</th>
<th>COST AND EFFICACY DETAILS CONSIDERED</th>
<th>MAIN RESULTS</th>
</tr>
</thead>
</table>
| Insinga et al., 2002 United States | - Cost-effectiveness of MS/MS-based neonatal screening of MCADD only, then of MCADD together with certain fatty acid oxidation disorders (LCHADD, VLCADD, SCADD, CPTII, GAII) and organic acidemias (GAI, PA, MMA, IVA, 3-MCC, β-KT and HMG)  
  - Sequential analysis by modelling with sensitivity analysis (base-case scenario based on very conservative assumptions regarding incidence, costs and health outcomes, and a more realistic scenario)  
  - Hypothetical cohort of 100,000 infants  
  - Cost-effectiveness threshold set at $50,000/QALY  
  - Societal perspective over the lifetime of the hypothetical cohort  
  - Probability of death: 16% for unscreened children with MCADD  
  - Probability of neurological impairment among survivors: about 15% | - The incremental costs, expressed in 2001 US dollars, include the equipment, consumables, staff, overhead expenses, laboratory costs, confirmatory tests, carnitine supplements and follow-up costs up to the age of 18 years, as well as the costs related to neurological impairment.  
  - Discount rate: 3%.  
  - Five health outcomes: asymptomatic; acute complications; mild neurological impairments; severe neurological impairments; death.  
  - N.B.: As mentioned by Venditti and colleagues [2003], the types of costs used here do not reflect a societal perspective. | - Under conservative assumptions, MCADD screening alone yields an ICER of $41,862/QALY. Under more realistic assumptions: ICER = $6,008/QALY.  
  - Adding the 12 other inborn errors of metabolism tends to yield a cost-effectiveness ratio within accepted norms for cost-effectiveness ($15,252/QALY). |

Abbreviations: QALY: quality-adjusted life-year; β-KT: beta-ketothiolase deficiency; CAH: congenital adrenal hyperplasia; CPTII: carnitine palmitoyltransferase deficiency type II; GAI: glutaric acidemia type I; GAII: glutaric acidemia type II; HMG: 3-hydroxy-3-methylglutaryl-CoA lyase deficiency; ICER: incremental cost-effectiveness ratio; IVA: isovaleric acidemia; LCHADD: long-chain hydroxyacyl-CoA dehydrogenase deficiency; MCADD: medium-chain acyl-CoA dehydrogenase deficiency; 3-MCC: 3-methylcrotonyl-CoA carboxylase deficiency; MMA: methylmalonic acidemia; MS/MS: tandem mass spectrometry; PKU: phenylketonuria; SCADD: short-chain acyl-CoA dehydrogenase deficiency; VLCADD: very-long-chain acyl-CoA dehydrogenase deficiency.
TABLE K-1

<table>
<thead>
<tr>
<th>AUTHORS, COUNTRY</th>
<th>DESCRIPTION OF STUDY AND METHODOLOGY</th>
<th>COST AND EFFICACY DETAILS CONSIDERED</th>
<th>MAIN RESULTS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Schoen et al., 2002 United States</td>
<td>Estimation of the costs and potential benefits of routine MS/MS-based neonatal screening of PKU, MCADD, MSUD, MMA and PA</td>
<td>Data on 32,000 newborns from the Kaiser Permanente Medical Care HMO (Northern California) and other published data.</td>
<td>Least favourable scenario (laboratory cost = $20/test): cost/QALY = $11,419</td>
</tr>
<tr>
<td></td>
<td>Base-case scenario and the most favourable and least favourable scenarios (higher laboratory costs, more false-positive test results, and smaller effects on mortality and morbidity rates)</td>
<td>Costs (2001 US dollars): sample collection and laboratory (including false positives), treatment, hospitalization, disabilities and special diets.</td>
<td>Most favourable scenario (laboratory cost = $7 per test), cost per QALY = $736. In the least favourable scenario, the cost of false positives is higher: mean per-test cost of false-positives = $12.25.</td>
</tr>
<tr>
<td></td>
<td>Payer perspective</td>
<td>Discount rate: 3%.</td>
<td>Base-case scenario (laboratory cost = $15/test): cost/QALY = $5,827.</td>
</tr>
<tr>
<td></td>
<td>Probability of death for unscreened infants with MCADD: 2.5%</td>
<td>Data on 32,000 newborns from the Kaiser Permanente Medical Care HMO (Northern California) and other published data.</td>
<td>Cost/QALY = $5,827.</td>
</tr>
<tr>
<td></td>
<td>Probability of neurological disabilities in survivors: 0%</td>
<td>Costs (2001 US dollars): sample collection and laboratory (including false positives), treatment, hospitalization, disabilities and special diets.</td>
<td>Discount rate: 3%.</td>
</tr>
</tbody>
</table>

Abbreviations: QALY: quality-adjusted life-year; HMO: health maintenance organization; MCADD: medium-chain acyl-CoA dehydrogenase deficiency; MMA: methylmalonic acidemia; MS/MS: tandem mass spectrometry; MSUD: maple syrup urine disease; PA: propionic acidemia; PKU: phenylketonuria.
### Economic aspects: cost estimates

**TABLE L-1**

<table>
<thead>
<tr>
<th>REAGENT*</th>
<th>APPROXIMATE COST PER SAMPLE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Internal standards (isotopes): 100 µL per sample (CA$800 per 500 L)</td>
<td>$0.01</td>
</tr>
<tr>
<td>Methanol: 100 µL per sample (CA$100 per 4 L)</td>
<td>$0.01</td>
</tr>
<tr>
<td>Butanol: 70 µL per sample (CA$60 per 1 L HPLC-grade)</td>
<td>$0.01</td>
</tr>
<tr>
<td>Aluminium</td>
<td>$0.01</td>
</tr>
<tr>
<td>Acetonitrile: 70 µL per sample (CA$350 per 4 L)</td>
<td>$0.09</td>
</tr>
<tr>
<td>Three 96-well plates† (CA$400 per 100 plates)</td>
<td>$0.20</td>
</tr>
<tr>
<td>Multichannel pipettor tips</td>
<td>$0.02</td>
</tr>
<tr>
<td><strong>Total per sample</strong></td>
<td><strong>$0.35</strong></td>
</tr>
</tbody>
</table>

* This information was provided by a clinical biochemist specialized in genetics and was confirmed by other experts in the field. The costs of the pipettes, vials, punches and other instruments used in all types of screening were not taken into account.
† The third plate is used for succinylacetone assays.

**TABLE L-2**

<table>
<thead>
<tr>
<th>CATEGORY OF PROFESSIONAL</th>
<th>HOURLY RATE</th>
<th>TASKS IN MS/MS-BASED SCREENING</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nurse* (hourly rate varies from $23 to $42, including benefits)</td>
<td>$42/hour</td>
<td>Sample collection: cost is the same as for current method and was therefore not factored into the calculation (± 15 minutes)</td>
</tr>
<tr>
<td>Computer technician†</td>
<td>$30/hour</td>
<td>Data entry: cost similar to that of current method and therefore not considered in the calculation (amount of time not known)</td>
</tr>
<tr>
<td>Laboratory technician† (level 12)</td>
<td>$31/hour</td>
<td>Instrument preparation and operation, checking and monitoring of samples (two scenarios are considered: 1.5 FTEs and 2 FTEs, since screening has to operate on an annual basis of 365 days)</td>
</tr>
<tr>
<td>Clinical biochemist* (hourly rate varies from $37.99 to $53.19, including benefits)</td>
<td>$53/hour</td>
<td>Analysis: depending on the expert consulted, this stage of the process is the same as for the current screening method. The related costs were therefore not taken into account.</td>
</tr>
</tbody>
</table>

## APPENDIX M

### Summary table: issues associated with newborn screening programs

**TABLE M-1**

<table>
<thead>
<tr>
<th>ISSUES (MAIN STAKEHOLDERS)</th>
<th>ASPECTS</th>
<th>SELECTED REFERENCES</th>
</tr>
</thead>
</table>
| Impact of uncertainty (families) | Waiting time for diagnostic results  
Meaning of results: false positive, true positive, false negative  
Lack of knowledge regarding definitive manifestations  
| Impact of knowledge (families) | Reactions to an early diagnosis  
Effects on the parent-child relationship  
Influence of follow-up/treatment  
Repercussions on reproductive decision making | Tran *et al.*, 2006; Crone *et al.*, 2005; Green *et al.*, 2004; Pandor *et al.*, 2004; Campbell and Ross, 2003; Lloyd-Puryear and Forsman, 2002 |
| Stigmatisation (families) | Genetic discrimination by third parties  
Unexpected identification of carrier status  
Determination of mis-attributed parentage | Tran *et al.*, 2006; McCabe and McCabe, 2004; 2002; Oliver *et al.*, 2004; Read, 2004; Wilcken, 2003 |
| Consent (families and health care professionals) | Explicit and informed consent  
Uptake of screening  
Timing of patient education  
Professional roles regarding the provision of patient education | Faulkner *et al.*, 2006; Feuchtbaurm *et al.*, 2006b; Grosse *et al.*, 2006a; Newson, 2006; Hargreaves *et al.*, 2005; Green *et al.*, 2004; McCabe and McCabe, 2004; Kim *et al.*, 2003; Clague and Thomas, 2002; Liebl *et al.*, 2002b; Lloyd-Puryear and Forsman, 2002 |
| Professional knowledge gap (health care professionals) | Professional education  
<table>
<thead>
<tr>
<th>ISSUES (MAIN STAKEHOLDERS)</th>
<th>ASPECTS</th>
<th>SELECTED REFERENCES</th>
</tr>
</thead>
</table>
| Societal impacts (society) | Accessibility and equity  
Legal issues  
Societal roles in health care planning and research | Alexander and van Dyck, 2006; ASTHO, 2005; Avard, 2005; Waisbren et al., 2004; Al-Odaib et al., 2003; Campbell and Ross, 2003 |
| Research and evaluation (health care system and society) | Lack of evidence-based data*  
Provision of services  
Storage of samples | Cunningham et al., 2005; Seashore and Seashore, 2005; McCabe and McCabe, 2004; Khoury et al., 2002; Farrell et al., 2001 |
| Organisational aspects (health care system) | Technical: maintenance and quality assurance  
Professional: availability, interdependence and expertise  
Integrated system: development and sustainability | Feuchtbauem et al., 2006a; Frazier et al., 2006; Marsden et al., 2006; ASTHO, 2005; Carlson, 2004; Comeau et al., 2004; Wilcken, 2004; Kim et al., 2003; Roschinger et al., 2003; Therrell, 2003; Wilcken, 2003; Wiley et al., 2003; Elliman et al., 2002; Liebl et al., 2002a; McCabe and McCabe, 2002; McCabe et al., 2002 |

* There is a need to accumulate more evidence on the efficacy of new treatments, the risk of diseases by ethnic group and the penetrance of the mutations.
# Proposed decisional scenarios

## Table N-1

<table>
<thead>
<tr>
<th>OPTION</th>
<th>MODALITIES</th>
<th>RATIONALE</th>
<th>ADVANTAGES</th>
<th>DISADVANTAGES</th>
<th>ISSUES</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Pilot study</strong></td>
<td>a) Preimplementation pilot phase:</td>
<td>Bridging gaps in knowledge and cost estimation</td>
<td><strong>Modalities a and b:</strong></td>
<td><strong>Modalities a and b:</strong></td>
<td><strong>Modalities a and b:</strong></td>
</tr>
<tr>
<td></td>
<td>• Neonatal screening of PKU and TT1: MS/MS and current methods</td>
<td></td>
<td>• Trial run of the MS/MS-based neonatal screening organization</td>
<td>• Costs associated with the evaluation</td>
<td>Necessity of:</td>
</tr>
<tr>
<td></td>
<td>• Neonatal MCADD screening: MS/MS</td>
<td></td>
<td>• Field evaluation of the costs of current and MS/MS-based screening methods</td>
<td>• Need to maintain current neonatal screening methods for PKU and TT1</td>
<td>clear objectives</td>
</tr>
<tr>
<td></td>
<td>b) Comparative pilot study:</td>
<td></td>
<td>• Comparison of performance of MS/MS and current methods of screening: PKU</td>
<td>• Possibility of disrupting the neonatal screening process</td>
<td>rigorous study design</td>
</tr>
<tr>
<td></td>
<td>• Experimental group: MS/MS-based neonatal screening of PKU, TT1 and MCADD</td>
<td></td>
<td>and TT1</td>
<td>• Limited advancement of clinical knowledge on MCADD</td>
<td>meticulous planning</td>
</tr>
<tr>
<td></td>
<td>• Control group: current PKU and TT1 screening methods; no neonatal screening for MCADD</td>
<td></td>
<td>• Advancement of genetic and epidemiological knowledge on MCADD</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• Collection of short term clinical data on MCADD.</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td><strong>Modality b:</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• Possibility of evaluating the benefits of neonatal MCADD screening</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td><strong>Modality a:</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• Impossible to assess the benefits of neonatal MCADD screening</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td><strong>Modality b:</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• Difficulty identifying a control group</td>
<td></td>
<td></td>
</tr>
<tr>
<td>OPTION</td>
<td>MODALITIES</td>
<td>RATIONALE</td>
<td>ADVANTAGES</td>
<td>DISADVANTAGES</td>
<td>ISSUES</td>
</tr>
<tr>
<td>------------------------------------------</td>
<td>------------</td>
<td>---------------------------------------------------------------------------</td>
<td>---------------------------------------------------------------------------</td>
<td>--------------------------------------------------------------------------------</td>
<td>---------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td><strong>Postponed technology transfer:</strong> PKU, TT1 and MCADD</td>
<td>Does not apply</td>
<td>Complete the validation studies of MS/MS-based succinylacetone quantification for neonatal TT1 screening in a common protocol for all metabolites</td>
<td>▪ Concurrent, hence simpler, technology transfer for these three diseases</td>
<td>Delay in offering neonatal MCADD screening</td>
<td>Dealing with pressures in favour of technology transfer</td>
</tr>
</tbody>
</table>
| **Gradual technology transfer:**          | For neonatal TT1 screening | ▪ Immediate: a) Maintain current methods for tyrosine and succinylacetone quantification  
   b) Maintain current methods for succinylacetone quantification and use MS/MS for tyrosine quantification  
   c) Use a separate MS/MS protocol for succinylacetone quantification:  
   ▪ by alternately using the same instrument*  
   ▪ or by using two instruments  
   ▪ Immediate provision of neonatal MCADD screening with the investment in the technology being cost-effective from the outset given the use of MS/MS for neonatal PKU screening | ▪ Need to maintain all of the current methods of neonatal TT1 screening alongside MS/MS  
   ▪ Need to maintain the current methods of succinylacetone quantification alongside MS/MS  
   ▪ Costs (investment and/or instrument lifespan) potentially higher with the use of a separate MS/MS succinylacetone quantification | ▪ Coordinating blood sample management to ensure smooth operation of neonatal screening for all diseases, including congenital hypothyroidism.  
   ▪ Managing organizational problems and avoiding delays in communicating results.  
   ▪ Identifying and resolving a priori any problems concerning the additional transitional step required if the results of the validation of the protocol common to all metabolites are conclusive. | ▪ Determining the best organizational strategy (use of one or two MS/MS instruments). |


McCandless SE. A primer on expanded newborn screening by tandem mass spectrometry. Prim Care 2004;31(3):583-604, ix-x.


Wilcken B. Screening of newborns for metabolic disorders with mass spectrometry. JAMA 2004;291(12):1444; author reply:5.


